STATISTICAL ANALYSIS OF SPATIAL PATTERN OF MALARIA: THE CASE OF WESTERN WOLLEGA IN GIMBI ZONE, ETHIOPIA

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS
COLLEGE OF NATURAL AND COMPUTITIONAL SCIENCES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTERS OF SCIENCE IN STATISTICS
(SPECIALIZATION: BIOSTATISTICS)

By

TEMESGEN SENBETO

ADVISOR: DEJEN TESFAW (PhD)

AUGUST, 2015

GONDAR, ETHIOPIA

University of Gondar
School of Graduate Studies
Department of Statistics

Title of Thesis: Statistical analysis of spatial pattern of malaria: The case of Western Wollega in Gimbi zone, Oromia, Ethiopia.

By
Temesgen Senbeto Wolde

Approved by the Board of Examiners:

Name of Chairman	Signature	/ / Date
Dejen Tesfaw (PhD) Name of Main Advisor	 Signature	14/08/2015 Date
Name of Co-Advisor	Signature	/ / Date
Name of Internal Examiner	Signature	/ / Date
Name of External Examiner	Signature	/ / Date

DECLARATION

As researcher of the thesis, I, the undersigned, assert that the thesis is my original work, has not been presented for degrees in any other University and all sources of materials used for the thesis have been duly acknowledged.

Temesgen Senbeto Wolde		/ /
Name	Signature	Date

Table of Contents

List of Tables

List of Figures

Acronyms

CSA	Central Statistical Agency
FHOM	Federal Ministry of Health
GDP	Growth Domestic Product
GFATM	Global Fund Aids, Tuberculosis and Malaria
IRS	Indoor residual insecticide spraying
ITN	Insecticide treated nets
ITNs	Insecticides-Treated Nets
LLINs	Long-lasting insecticide treated nets
LLITNs)	Long Lasting insecticide Treated Nets
LM	Lagrange Multiplier
NISR	National Institute of Statistics of Rwanda
NMMTSP	National Malaria Medium Term Strategic Plan
NMMTSP	National Malaria Medium Term Strategic Plan
OLS	Ordinary Least Squares
PMI	President's Malaria Initiative
RLM	Robust Langrage Multiplier
UNICEF	United Nations Children's Funds
VIF	Variance inflation factor
WHO	World Health Organization

Acknowledgments

First, and foremost, I express deepest gratitude to **Dr**. **Dejen Tesfaw,** my advisor for his unlimited constructive advice, suggestions, ideas and comments. My thanks also to Central Statistical Agency and Western Wollega Zone Health Bureau for providing me with the data for this thesis.

Furthermore, I want to thank all my friends, my family members and others who have provided me a great help and moral support directly and indirectly contributed to wards the completion of this study.

Grateful acknowledgement is also expressed to the Department of Statistics, University of Gondar for kind assistance in many ways.

Abstract

Malaria is a serious health threat in the World, mostly in Africa, where it has been estimated that 90% of the world's cases occur. It is the major cause of health problems in Ethiopia, accounting for more than thousands of cases and deaths occurring annually. The risks of morbidity and mortality associated with malaria incidences are characterized by spatial variations across the country. The main objective of this study was to analyze spatial patterns of malaria distribution in Western Wollega Zone, Oromia Region, Ethiopia. Malaria incidence data for 2014 from all health centers of the zone was obtained from Gimbi Hospital, population size obtained from Central Statistical Agency and meteorological data were obtained from Gimbi Agricultural Bureau. The statistical methods used in this study include global and local measures of spatial autocorrelation as well as spatial autoregressive model. The results of the study indicated that malaria incidence varies according to geographical location, with eco-climatic condition and showed significant positive spatial autocorrelation. Significant local clustering of malaria incidence occurred between pairs of neighboring Woredas. Global Moran's *I*, Geary's *C* and Moran scatter plot are used in determining distribution of malaria incidence whereas the local Moran's *I* and Local Ord and Getis' Gi* statistic were used in identifying areas of hot spot and cold spot for giving strong care to monitor and reduce malaria incidence distribution. The values for Global Moran's *I* showed that the presence of significant malaria incidence clustering in Western Wollega Zone and in fifteen woredas significant malaria incidence clustering of similar values were observed by using cluster map while only in five woredas significant malaria incidence clustering of dissimilar values was observed. Malaria incidence was higher in the eastern part of the zone and lower in the northern part of the zone. The results of spatial lag model indicated that there were a statistically significant effect between malaria incidence and meteorological variables such as rainfall, maximum temperature minimum temperature, middle land and low land area.

CHAPTER ONE
INTRODUCTION

1.1. Background of the study

Malaria is a mosquito-borne infectious disease of humans and other animals caused by parasitic protozoans (a group of single-celled microorganism) belonging to the genus Plasmodium. Malaria causes symptoms that typically include fever, fatigue, vomiting and headaches. In severe cases it can cause yellow skins, seizure, and coma or death. The disease is transmitted by the biting of mosquitos, and the symptoms usually begin ten to fifteen days after being bitten. In those who have not been appropriately treated disease may recur months later. In those who have recently survived an infection, re-infection typically causes milder symptoms. This partial resistance disappears over months to years if there is no ongoing exposure to malaria (Mwangangi *et al*., 2011).

The disease is transmitted most commonly by an infected female Anopheles mosquito. The mosquito bite introduces the parasites from the mosquito's saliva into a person's blood. The parasites then travel to the liver where they mature and reproduce. Five species of Plasmodium can infect and be spread by humans. Most deaths are caused by P. falciparum because P.vivax and P.malariae generally cause a milder form of malaria. The species P.knowlesi rarely causes disease in humans. The transmission intensity is highly sensitive to environmental variations. Variations in transmission intensity have been observed within very small localities due to geographical, biological or socio-economic factors. Understanding the heterogeneity in transmission and human exposure to malaria infection is critical to optimize control programs and targeting interventions. Anopheles mosquito suitable habitat is allocation where biophysical conditions are adequate for its life cycle. Habitat is the location or environment where the organism is most likely to be naturally found (Planting *et al*., 2014).

Malaria infection is the invasion of host organism’s bodily tissues by the disease causing organisms, their multiplication and the reaction of host tissues to these organisms and the toxins they produce in blood cells. As the malaria parasites enter the blood stream they infect and destroy red blood cells. Destruction of these essential cells leads to fever and

flu-like symptoms, such as chills, headache, muscle aches, tiredness, nausea, vomiting and diarrhea. These initial symptoms are non-specific. In other words, they are self-reported symptoms that do not indicate a specific disease process (Auto *et al*., 2013).

The risk of disease can be reduced by preventing mosquito bites by using Mosquito nets and insect repellents or with mosquito-control measures such as spraying insecticides and draining standing water. Several medications are available to prevent malaria in travellers to areas where the disease is common. Occasional doses of the medication sulfadoxine or pyrimethamine are recommended in infants and after the first trimester of pregnancy in areas with high rates of malaria. Despite a need, no effective vaccine exists, although efforts to develop one are ongoing. The recommended treatment for malaria is a combination of antimalarial medications that includes an artemisinin. The second medication may be either mefloquine or sulfadoxine/pyrimethamine. Quinine along with doxycycline may be used if an artemisinin is not available (Karema *et al*., 2012).

In the global context, malaria remains one of the main global health problems of our time, causing an estimated 247 million people that led to nearly 881,000 deaths. Approximately 86% of world's estimated cases and 91% of deaths occur in Africa (WHO, 2010). Malaria is the leading killer of children in Africa, accounting for approximately 20% of all-cause mortality in children under the age of five (Abdullah, 2010).

Globally, malaria not only has a toll on health, it negatively impacts economic development resulting in malaria-endemic countries having lower rates of economic growth. Especially in Africa, where malaria accounts for 30% to 50% of hospital admissions and up to 50% of outpatient visits in high-transmission areas. The World Bank estimates that malaria alone slows African economies by 1.3% per year a 32% reduction in African GDP over 35 years. Malaria costs African economies US $12 billion annually (Sachs and Mananey, 2011).

According to (WHO, 2010), about 81% of all malaria cases and 91% of all malaria related deaths occurred in the Africa region. There are 43 malaria-endemic countries in the Africa region. Malaria infected an estimated 216 million people and killed 655,000

people, most of whom were children under the age of five in sub-Saharan Africa. Despite the current burden of disease, malaria is preventable and treatable. Congress has increasingly recognized malaria as an important foreign policy issue and the United States has become a major player in the global response to the disease.

The National malaria control program in Ethiopia assisted by several international programs (such as Global Fund which has been fighting AIDS, Tuberculosis and Malaria (GFATM), World Health Organization (WHO, 2012), the President's Malaria Initiative (PMI), United Nations Children's Funds (UNICEF, World Bank, etc.) adopted several key strategies for malaria control including increasing the coverage for anti-malarial treatment, long-lasting insecticide treated nets (LLINs) and indoor residual insecticide spraying (IRS).

The President's Malaria Initiative (2010) of Ethiopia, national malaria plan indicated that malaria is ranked as the leading communicable disease in Ethiopia, accounting for about 30% of the overall disability adjusted life years lost. Approximately, 75% of the country is malarious with about 68% of the total population living in areas at risk of malaria.

According to Ethiopia's Federal Ministry of Health (FMOH, 2009), malaria was the first cause of outpatient visits, health facility admissions and in-patient deaths, accounting for 12% of out-patient visits and 9.9% of admissions. However, as 36% of the population does not have access to health care services, these figures probably under-represent the true burden of malaria in the country. Increasing the understanding of the distribution dynamics of malaria and their relationship could suggest improvements for malaria control efforts.

Malaria is the leading cause of morbidity and mortality in Ethiopia, accounting for over nine million cases and thousands of deaths annually. The risks of morbidity and mortality associated with malaria are characterized by spatial variation across the country. Consequently, we recognize the spatial variation of malaria by means of spatial autocorrelation.

Spatial autocorrelation can be expressed as the relationship among values of a single variable that comes from the geographic arrangement of the areas in which these values

occur. It measures the similarity of objects within an area; the degree to which a spatial phenomenon is correlated to itself in space (Cliff and Ord, 1981), the level of interdependence between the variables, the nature and strength of the interdependence. Spatial autocorrelation is an assessment of the correlation of a variable in reference to spatial location of the variable.

Spatial autocorrelation statistics such as global Moran's *I* and Geary's *C* are used to estimate the overall degree of spatial autocorrelation for a dataset. The possibility of spatial heterogeneity suggested that the estimated degree of autocorrelation may vary significantly across geographic space, whereas, local spatial autocorrelation statistics provided estimates disaggregated to the level of the spatial analysis units, allowing assessment of the dependency relationships across space. *G*ets and Ord statistics used to compare neighborhoods to a global average and regions of strong autocorrelation. Global spatial autocorrelation statistics such as the global identified local Moran's *I* and Geary's *C* described the overall spatial dependence of malaria over the entire region whereas local spatial autocorrelation statistics such as the local Moran's *I* (Anselin,1995) and Gets and Ord Gi* (Getis and Ord ,1992) are useful to identify local patterns.

The main goal of this study is to examine spatial patterns of malaria distribution using district level malaria incidence data. It seeks to identify malaria "hotspot" Woredas by producing map of clustering observation and fit appropriate spatial models for malaria distribution in western wollega Gimbi zone, Oromia region, Ethiopia.

1.2. Statement of the Problem

Malaria cases depends on the environmental, seasonal, climatic and others different socioeconomic factors. Demographic, eco-climatic mortality factors, age and sex vary by geographical location and many authors recommended that targeting interventions to the high malaria case are omitted due to inconsideration of spatial dependence. According to Smith (2003) regions that are in closer proximity are expected to have similar malaria cases because of similar eco- climatic situation and demographic characteristics. Spatial models explain malaria morbidity variation by geographical location better than non-spatial models when limited data is available for meteorological variables. Spatial model

is devoted to measure neighboring influence and involved in different area of research. Cliff and Ord (1973) revealed that the analysis of spatially located data is one of the basic concerns of the statistician and so it becomes increasingly important in other fields of study. Malaria incidences, which also vary spatially, raise the need for spatial models for covariates. Most of the time malaria incidence was caused from stored water for a long time. In this study, the spatial distributions of malaria was assessed using spatial model along with meteorological and environmental variables of malaria incidence in western wollega Gimbi zone and identified whether the distribution of malaria is clustered or not. Controlling malaria at zonal level would need identifications of climatic factors related to malaria. Hence, considering seasonal and geographical variations malaria transmission has been difficult due to a lack of resource and time as well as usage of inappropriate statistical methods and data. Therefore, modeling has suggestions for malaria clear control and jeopardy management. Environmental variation jeopardies can be counted using spatial models of malaria and disease heterogeneity. Therefore, this study concentrated on the number of affected people by malaria incidence and would attempt to address the following problems:

- What are the major significant incidence factors of malaria?
- What is the implication of observing significant spatial dependencies in the data?
- How to determine the effect of variables on malaria distributions in the study area?
- In what area there is a hot spot of malaria incidence?

1.3. Objective of the study

General objective:

The general objective of the study is to identify the most important factors that affect the spatial distribution of malaria incidence in the case of Western Wollega Zone.

The specific objectives:

- To discuss and measure the intensity of spatial autocorrelation in the distributions of malaria incidence.
- To characterize the distribution of malaria incidence.

1.4. Significance of the study

- Analysis of malaria incidence was important to address and identify the significant incidence factors of malaria transmission at study areas, to take remedial actions in order to control the patterns of malaria transmission periodically and geographically as well as to recommend the spatial distributions of malaria to policy makers and stakeholder. In addition to this, this study is used to classify the woreda of the zone into high and low risk groups so as to give information on how to optimize available resources for malaria control.
- The results of this study can provide information to governments and concerned bodies in setting policies, strategies and further investigation for reduction of malaria incedence.

CHAPTER TWO

LITERATURE REVIEW

Many authors in various disciplines discussed geographical distribution of diseases as a key element in epidemiologic research, depending on importance given to the description of health events such as patients, place and time. Researchers have been focusing on the relationship between demographic factors and health that extremely determine geographical distribution of diseases. The description of spatial patterns of disease incidence and mortality can be defined as geographical epidemiology.

The proposed modeling approach appropriately accounts for spatial and temporal dependence typical in studies of infectious diseases such as malaria. Results demonstrate that the proposed modeling approach is robust and can be useful in understanding the impact of climate change on the spread of malaria. Additionally, the model can be applied to analyses the spread of other infectious diseases and in optimizing management efforts (e.g., drug policy changes) on the spread of malaria. With a more rigorous effort, this modeling framework can be extended to account for socio-economic factors as well as other important factors such as access to health, information on drug policy, and drug resistance (Arab *et al.*, 2014).

Grilletet *et al.* (2010) used local spatial statistics and geographically weighted regression to determine the spatial pattern of malaria incidence and persistence in northeastern Venezuela. It was reported that the geographical weighted regression model greatly improved predictions of malaria risk compared with multiple linear regression models. Results also indicated that disease persistence was associated with greater human population density, lower elevations, and proximity to aquatic habitats.

Matthew (2013) made a study in Harris County, Texas. The main purpose of the study was to examine the spatial distribution of malaria cases in Harris County during the period of 2010 to 2012 using GIS software. Spatial analytical techniques mainly Global Moran's *I*, Geary's *C* and Local Indicators of spatial autocorrelation were applied. Determining the distribution of malaria was based on the incidence rate and intensity measure after checking the assumption of complete spatial randomness. The result of the

study indicated that from nine different areas studied for the presence of malaria clustering only two were found to have significant clusters and high malaria incidence rate.

The combination of different malaria control methods has been proved efficient (NISR, 2012; Presidents Malaria initiative, 2013). In between 2006 and 2012 these technologies contributed to the reduction of malaria microscopically confirmed cases by 72% for all ages and 82% for children below five years of age. Malaria death decreased by 47% for all ages and 77% for children below five years (Kerama *et al.*, 2012).

The biophysical (climatic and topographic) that can determine the regions with high endemicity have been objects of different researchers (Zayeri *et al*., 2011). Anopheles mosquito proliferation depends on environmental factors like temperature, rainfall and humidity in association with vegetation cover and hydrology, especially water bodies. Altitude is also an important factors and Anopheles mosquito prefers low altitude areas not only because they are characterized by high temperature and humidity especially in tropical regions but also because of their ability to retain water during and after rainy seasons (Fanello *et al.*, 2007).

Yeshiwondimet *et al*. (2009) examined the global and local patterns of malaria distribution in 543 villages in Ethiopia using individual-level morbidity data collected from six laboratory and treatment centers. It was reported that malaria incidence varies according to gender and age with age less five years and above showing a statistically significant malaria incidence. It was also observed that local clustering of malaria incidence between pairs of villages within distance lags were significant. Furthermore, malaria hot spots were displayed as risk maps that are useful for monitoring and spatial targeting of prevention and control measures.

Modeling of malaria helps to describe the existing spatial patterns of the disease, to understand it's causing factors especially the ecology of the vectors and to predict the future (Steven and Pfiffer, 2011). In their review they suggested that an adequate modeling of malaria must integrate the spatial and the classical statistics approach. The spatial modeling of malaria based on environmental variables such as temperature,

rainfall, humidity and topographic variables especially altitude that determine anopheles mosquito habitat has been applied indifferent parts of the world (Machaut *et al*, .2011).

The goal of the National Malaria Medium Term Strategic Plan (NMMTSP) 2008– 2013 is to reduce the occurrence of malaria in the country by 80 percent. This goal is in line with the Global initiative, that advocates a rapid scaling of interventions to achieve the roll back malaria target of universal coverage of 80 percent by 2010 and the Millennium Development Goals by 2015.

Awash *et al*. (2009) made a population based cohort study comprising 8,088 malaria cases in Adama, Ethiopia. The study was mainly designed to describe temporal and spatial clustering of malaria cases and to identify factors associated with malaria clustering. One result of the study indicates the existence of stable temporal and spatial malaria clustering in Adama. Global Moran's *I* was used. Another result of the study indicates that among all factors associated with malaria incidence maximum and minimum daily temperatures were the most possible reason for malaria clustering.

From the last two decades, the use of Insecticides-Treated Nets (ITNs) and Long Lasting insecticide Treated Nets (LLITNs) has been the most efficient malaria control measures in Rwanda. It has been incorporated in different vision and plans. The NISR (2012) suggested that in 2010, 80% of households had at least one LLIN or ITN. The nets ownership was the highest in the eastern province (90% of households).

Asnakew *et al*. (2012) analyzed malaria clustering in East Wollega, Ethiopia. In the study global spatial autocorrelation and local spatial autocorrelation were used to identify the patterns of malaria distribution in 410 villages. Statistical spatial analysis of malaria incidence by age, temperature and village through time revealed the presence of significant spatio-temporal variations. The result of local spatial statistics showed the presence of malaria clustering or hot spots in most villages. Malaria hot spots were identified by using cluster map that are useful for monitoring and targeting of prevention and control measures against the disease.

Tsai *et al.* (2009) employed spatial autocorrelation methodologies, including Global Moran's *I* and Local Getis-Ord statistics. The results indicated that cluster mapping helps

to illuminate issues such as the spatial aspects of correlations for leading health care events.

Lindsay *et al.* (2013) identified a reduction in the infection of malaria in Tanzania as associated with rainfall. It was found out that heavy rainfall may have flushed out anopheles mosquitoes from their breeding sites thereby increasing the mosquito population and showed a positive association between the abundance of anopheles mosquitoes and rainfall.

Malaria disease accounts for more than 44% of reported outpatient visits and an estimated 22% of under-five mortality in Ghana. Reported malaria cases represent only a small proportion of the actual number of episodes as majority of people with symptomatic infections are treated at home and are, therefore, not reported (WHO, 2010).

Mbogo *et al*. (2013) studied the seasonal dynamics and spatial distributions of Anopheles mosquitoes and Plasmodium falciparum parasites along the coast of Kenya. Using hand-held GPS, they recorded latitude and longitude data at each site, and produced the spatial distribution maps for three Anopheles species.

Yan (2013) presented a spatially distributed mosquito habitat modeling approach, integrating a Bayesian modeling method with Ecological Niche Factor Analysis using GIS. He used data for seven environmental variables to represent the environmental conditions of larval habitats in the Kenya highlands. Zhou *et al*. (2011) used GIS layers of larval habitats, land use type, human population distribution, house structure, and hydrologic schemes, overlaid with adult mosquito abundance, to investigate the impact of environmental heterogeneity and larval habitats on the spatial distribution of adult Anopheles mosquitoes in Uganda.

Mmbando *et al*. (2011) conducted a study of four cross-sectional malaria surveys in 14 villages located in highland, lowland, and urban areas of northeastern Tanzania during the rainy seasons. Their results showed a significant spatial variation of P. falciparum infection in the region, identifying altitude, socio-economic status, high bed net coverage, and urbanization as important factors associated with the spatial variability in malaria.

The Malaria Atlas Project developed the science of malaria cartography by modeling the global spatial distribution of P. falciparum malaria endemicity and focused on the spatial heterogeneity of malaria transmission intensity and effectively produced and used maps as essential tools for malaria control (Hay *et al*., 2011).

Topography has a great influence on mosquito replication and affects the rate of malaria cases. Higher topographies results in cooler temperatures, which limits the reproduction rate of the parasite. Entomologic studies in eight villages to investigate the patterns of malaria transmission in different ecologic zones in Eritrea showed a positive relationship between the malaria cases and topography. Mosquito collections conducted for 24 months showed that the biting rates in the higher elevations as a result of the lower temperatures were twice as high as the lowlands. The complexity of topography and landscape in the highlands contributes to the spatial heterogeneity of vector abundance and malaria transmission intensity. It has implications for the survival of the vector for different altitudes (Shillu, 2003).

Broker et al. (2002) used epidemiological and population data to see the spatial distributions of Helminthes (one type of parasites) in Cameroon. They used a Logistic regression model to identify significant environmental variables, which affect the transmission of infection. The variables used in the regression analysis were minimum and maximum land surface temperature. The result revealed that maximum temperature was an important variable in determining Helminthes distribution. At higher temperatures, it is realized that female adult mosquitoes feed more frequently and digest blood more rapidly.

CHAPTER THREE
DATA AND METHODOLOGY

3.1 Source of Data

This study is mainly based on secondary data that contain malaria case which is obtained from all Woreda health centers and hospitals of western Wollega, Gimbi zone under the Oromia Control and Prevention Bureau in 2014 G.C. The total numbers of malaria cases are 66,633 with overall malaria incidence rate of 40.69 per 1000 in the year of 2014. Here malaria incidence rate can be obtained by dividing malaria case for population projection. Population projection is obtained from Central Statistical Agency.

The population projection figures are based on the results of National Population and Housing Census of Ethiopia conducted in May 2007. The base population for the projection was obtained from the 2007 Population and Housing Census for each of the regions and adjusted to the mid of the census year, 1 July 2007. Up to now population projection figures at woreda and zonal levels are prepared and printed on yearly Statistical Abstract of the Agency (EDHS, 2011). But these projections were done for each of the regions based on the mathematical method and then the figures for the woreda and zonal levels were prepared by the ratio method.

In this study, the results of microscopic examination is recorded that include all the malaria cases such as P. falciparum and P.vivax which are recorded by age and sex admitted by malaria and clinical cases.

3.2. Variables included in the study

This study depends on meteorological variables that are important to malaria incidence distributions. The dependent variable is malaria incidence whereas the independent variables include average annual rainfall, average annual maximum temperature, average annual minimum temperature, and percentage of highland areas, percentage of midland areas and percentage of lowland areas.

3.3. Study area

Western Wollega Gimbi Zone is one of the 18 Zones in Oromia Region, Ethiopia. It has 19 woredas and one administrative Town and 488 Kebeles with population projection of 1,637,663 in 2014. The zone covers 13436.5 thousand sq. km. The climate conditions can be classified as: highland area (2.5%), midland area (72.4%) and lowland area (19.1%). All 19 woreda and one administrative Town are covered in the study. These woredas are: Begi, Kundala, Babogambel, ManaSibu, Qiltukara, Jarso, Najo, BojiDirmaji, BojiChoqorsa, Guliso, Ayira, Yubdo, Ganji, Lalo Asab, Homa, Ana Gimbi, Haru, Sayonole, Gimbi Town and Nole kaba.

3.4. Methodology of the study

3.4.1. The Spatial Autocorrelation

Spatial autocorrelation analysis is a technique used to detect event patterns and measures the extent to which the occurrence of an event in areal unit contains or makes more probable to the occurrence of an event in neighboring areal unit. It is defined as the relation among values of a single variable that is attributable to the geographic arrangement of areal units on a map. Spatial autocorrelation is a measure of interdependence between values of a variable at different geographic locations and can be used to identify the degree of clustering (Goodchild, 1987).

Spatial autocorrelation is like temporal autocorrelation and more complicated because of temporal autocorrelation can only deals with one direction in which what happens at one time can be affected only what happened in the past. But spatial autocorrelation can potentially go in any direction; mean that what happens at any one point in time can be influenced by both the past and the future (Anselin, 1992).

Spatial autocorrelation is a measure of spatial dependency that quantifies the degree of spatial clustering or dispersion in the values of a variable measured across a set of locations (Odland, 1988, Lee and Kretzschmar, 1993). It is the correlation among values of a single variable strictly attributable to their relatively close locational positions on a two-dimensional (2-D) surface, introducing a deviation from the independent observations assumption of classical statistics.

Spatial autocorrelation exists because real-world phenomena are typified by orderliness, (map) pattern, and systematic concentration, rather than randomness. Tobler's first law of geography encapsulates this situation: ''everything is related to everything else, but near things are more related than distant things''.

There are two basic types of spatial autocorrelation statistics. These are Global and local measure of spatial autocorrelation. Global measures of spatial autocorrelation is used to identify whether the values of a variable exhibit a significant over all pattern of regional clustering, whereas local measures of spatial autocorrelation is used to identify the location of significant high and low value clusters. Global spatial Autocorrelation includes Moran's Index I and Geary's Coefficient *C*; and local spatial autocorrelation includes Local Moran's I and Gets and Ord Gi*. District boundaries were dissolved to Region boundaries in order to carryout spatial analysis by regions. The spatial distribution based on variables is carried out in order to determine the region with high occurrence of malaria incidence.

In this study, global and local measures of spatial autocorrelation will be used first to diagnose univariate spatial autocorrelation in the absence of covariates. Then, a standard regression model will be estimated and diagnostics test will be conducted to determine whether the covariates sufficiently model the spatial dependence in the dependent variable. If they do not, the spatial autoregressive model specification indicated by the diagnostic will be fitted.

3.4.2. Global and Local Measures of Spatial Autocorrelation

In spatial data analysis, it is necessary to determine whether or not identifiable spatial pattern exists. Here to test global spatial autocorrelation is to observe whether the data as a whole display spatial autocorrelation, the strength and direction of any spatial autocorrelation and while tests of local spatial autocorrelation is to identify particular observations that are auto correlated with neighboring observations of the dependent variable of interest and determine the strength. In addition to this based on the statistic, it is possible to identify the direction (positive or negative) of spatial autocorrelation (Anselin, 1995) in the case of local spatial autocorrelation.

3.4.2.1. Global Measures of Spatial Autocorrelation

Global clustering methods are used to test for spatial clustering throughout the study region without the ability of locating specific clusters sites. Their results provide a single statistic that measures the degree of spatial clustering, the statistical significance of which can also be assessed.

Spatial autocorrelation is an assessment of the correlation of a variable in reference to spatial location of the variable, which is a match between location similarities and attributes similarity. Moran's *I* is the more popular test statistic for spatial autocorrelation.

The Γ index (Anselin, 1992) is used for testing global and local spatial autocorrelation. The Γ index consists of the sum of the cross products of the corresponding elements W_{ij}, Y_{ij} of two matrices, W and Y with corresponding the row (i location) and the column (j location).

$$\Gamma = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} y_{ij} \qquad [1]$$

where W_{ij} is the ij^{th} element of spatial weights matrix W, Y_{ij} is the product of the two values $Y_{ij}=Y_iY_j$ or its squared difference $Y_{ij}= (Y_i-Y_j)^2$ of dependent variable at *i* and *j* locations, respectively and N is the number of observations. Measures of spatial autocorrelation are variants of this Γ index, with the Y_{ij} elements in Y reflecting how value is different or similar in the particular form of the Γ index.

3.5. Methods of Measuring Spatial Autocorrelation

3.5.1. Spatial Weights Matrix

A general spatial weight matrix can be defined as a symmetric binary contiguity matrix, which can be generated from topological information based on either adjacency or distance criteria. A fundamental characteristic distinguishing spatial data from time series data is the spatial arrangement of the observations. The spatial linkages or proximity of the observations are measured by defining a spatial weight matrix. The spatial weight matrix represents the strength of the potential interaction between locations. However, it

has to be noted that the determination of the proper specification for the elements of a spatial weight matrix is one of the difficult and controversial methodological issues in spatial data analysis (Odland, 1987).

To express the degree of proximity between observations in space we may attribute a value of one if the observations are nearby (neighbors) and zero otherwise. Spatial weight matrix can be explained through distance weight (Threshold Distance and k-nearest Neighbors) and Neighborhood/Contiguity weight.

According to Getis and Ord (1992) the administrative center of the observation units could adequately represent the location of the observation and the non-standardized spatial weight matrix *W (d)* for distance (*d)* is defined as:

$$w_{ij}(d) = \begin{cases} 1 & \text{if hypot } (i,j) \le d, i \ne j \\ 0 & \text{otherwise} \end{cases} \qquad [2]$$

where

$\text{hypot}(i,j) = \sqrt{(x_i - \bar{x})^2 (x_j - \bar{x})^2}$ x_i and x_j are values of dependent variables at i and j locations, respectively. The contiguity relation in terms of sets of neighbors of zones or sites having common boundary (Tobler, 1970), can be coded in the form of a spatial weight matrix W, with a zero diagonal, and the off- diagonal non-zero elements often scaled to sum to unity in each row which is standardized weights matrices and can be calculated as follows:

$$W_{ij} = \frac{w_{ij}}{\sum_{j=1}^{N} w_{ij}} \qquad [3]$$

In non-standardized binary spatial weight matrix the element of spatial weight matrix is 1 if location i is adjacent to location j and 0, otherwise and can be given as follows:

$$w_{ij} = \begin{cases} 1 & \text{if } i \text{ is linked to } j \\ 0 & \text{otherwise} \end{cases} \qquad [4]$$

Spatial weights matrices are row-standardized in which the sum of the weights for each row equals one. As a result, the spatial influence from neighbors is a weighted average of this influence across the neighbors. The definition of neighbors is a critical decision in the modeling of spatial autocorrelation. Closely related is the form and extent of spatial dependence between neighbors. The simplest definition of neighbors is the contiguity case. Contiguity analysis is an important method to assess unusual features in the connectivity distribution. The administrative woredas considered in this study are highly irregular on both shape and size. Neighborhood relations are defined as either Rooks case, Bishop's case, or Queen's (King's) case.

- **Rook's case** considers contiguity is by a neighborhood of four locations adjacent to each cell (locations which share a common border are considered as neighbors), as shown in Figure 1.
- **A bishop case** definition considers objects sharing a common vertex as neighbors (as shown in the Figure 2).
- **Queen's case** considers all neighborhoods (common boundary and or edge) to define neighborhood. It incorporates both the rook and bishop definitions as any object sharing either a common edge or vertex to be considered as a neighbor (as shown in the Figure 3).

Figure 1 **Figure2** **Figure3**

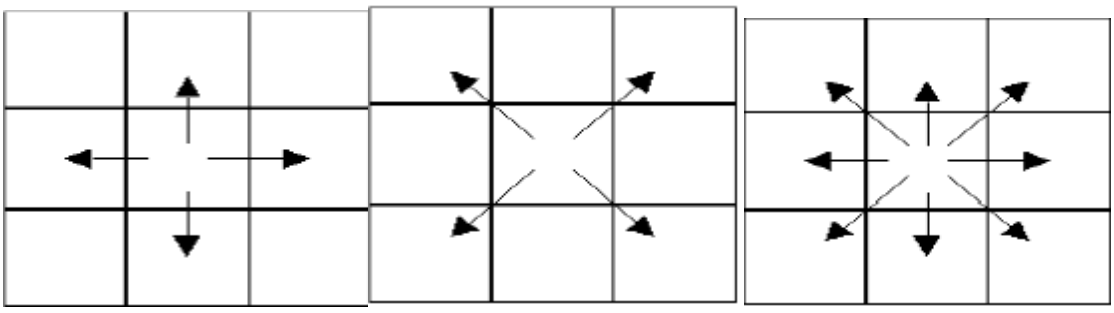


Figure 3.1: Contiguity case of representation of spatial weight matrix

3.5.2. Tests of Spatial Autocorrelation

The two most commonly used measures for spatial autocorrelation are Moran's *I* and Geary's *C* statistics. As we have mentioned above these test showed the degree of spatial

association as reflected in the dataset as a whole. The Moran's *I* is based on cross products to measure value association and the Geary's *C* employs squared differences (Anselin, 1992). Test for spatial autocorrelation are designed to quantify the extent of clustering and to allow for statistical inference. The null hypothesis under the normality and independence assumptions is given as follows:

H_o: No spatial autocorrelation ($\rho=0$) versus

H_1: There is spatial autocorrelation ($\rho\neq0$) or spatial dependence.

3.5.3. Global Moran's *I*

Global Moran's *I* is one of the Global autocorrelation that examines whether spatial correlation exists or not over an entire region, and it can be calculated as follow as:

$$I = \frac{N}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}} \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum(x_i - \bar{x})^2} \quad [5]$$

where N is the number of observations of the whole region, x_i and x_j are the observations at locations of i and j, $\bar{x}$ is the sample mean of x and w_{ij} is an element of spatial weights matrix of *w*. The selection of neighbors is formally specified in the weights matrix, which depicts the relationship between an element and its surrounding elements. In distance-based weight matrix, each distance class is specified as a threshold distance, such that all locations within the given distance are considered to be "neighbors" (the value not equal to zero in the matrix) in the distance-based weight matrix *W*. Usually, normal approximation global Moran's *I* can be standardized and calculated as follows:

$$Z(I) = \frac{I - E(I)}{\sqrt{\text{var}(I)}} \quad [6]$$

where,

$$E(I) = \frac{-1}{n-1}$$

The variance of Moran's *I* and Geary's *C* will vary under the assumptions normality and randomization. Under the normality assumption the variance of Moran's *I* Var $(I)_N$ is given as:

$$Var(I)_N = \frac{N^2(N-1)W_1 - N(N-1)W_2 - 2W_0}{(N+1)(N-1)W_0} \quad [7]$$

And under randomization Var $(I)_R$ is given by:

$$Var(I)_R = \frac{N(W_1(N^2-3N+3) - NW_2 + 3W_0{}^2) - K(W_1(N^2-N) - 2NW_2 + 6W_0{}^2}{(N-1)(N-2)(N-3)W_0{}^2} - \left(\frac{1}{N-1}\right)^2 \quad [8]$$

where,

$$W_0 = \sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}, i \neq j$$

$$W_1 = 0.5\sum_{i=1}^{N}\sum_{j=1}^{N} (W_{ij} + W_{ji})^2, i \neq j$$

$$W_2 = \sum_{i=1}^{N}\left(\sum_{j=1}^{N} W_{ij} + W_{ji}\right)^2$$

$$K = \frac{N\sum_{i=1}^{N}(x_i - \bar{x})^4}{\sum_{i=1}^{N}\left((x_i - \bar{x})^2\right)^2}$$

The value of Moran's *I* range from -1 to +1. Here a significant negative value shows that nearby locations tend to have different values (i.e. spatial dispersion), an insignificant value indicates that nearby locations tend to have random values, and a significant positive value indicates that nearby locations tend to have similar values (i.e. spatial

clustering). By testing for significant levels of positive global spatial autocorrelation, it is therefore possible to statistically identify the presence of malaria incidence in regional.

A positive global Moran's *I* that differs significantly from the expected value under the null hypothesis indicates positive spatial autocorrelation and implying the clustering of similar values (i.e, high values are found closer together, and low values are found closer together) on the dependent variable among neighboring observations. A negative global Moran's *I* that differs significantly from the expected value under the null hypothesis indicates negative spatial autocorrelation and implies the clustering of dissimilar values (means high values are found far away from other high values, and low values are found far away from other low values) on the dependent variable among neighboring observations (Anselin, 1992). The null hypothesis of no spatial autocorrelation will be rejected if the calculated value of $|Z(I)| \geq Z_{\alpha/2}$.

3.5.4. Global Geary's C

Global Geary's *C* depends on the difference between neighboring values of a variable. It is similar to the Durbin-Watson test. Global Geary's *C* explains the value of similarity or dissimilarity as the squared difference in values between neighboring observations. Geary's C interactions are not the cross product of the deviations from the mean, but the deviations in intensities of each observation location with one another**.** The Global Geary's *c* can be calculated by using the following formula:

$$C = \frac{N-1}{2W_0} \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}(x_i - x_j)^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \qquad [9]$$

Where, all notations are expressed in the equations of [5] and [8].

For $n \sim \infty$, Z(C) follows the standard normal distribution. That means $Z(C) \sim N(0,1)$ and the variable x follows an asymptotic normal distribution. Therefore, the Z-statistics of Geary C is given by:

$$Z(C) = \frac{C - E(C)}{sd(C)} \qquad [10]$$

Where, the standard deviation of the Geary's coefficient *C* is given by:

$$sd(C) = \sqrt{Var(C)}$$

But $E(C)_N = E(C)_R = 1$, where *N*=normal, *R*=random in this case, respectively.

The variance of Geary's *C* under normality $Var(C)_N$ and variance of the Geary's *C* under randomization $Var(C)_R$ are given as follows, respectively.

$$Var(C)_N = \frac{((2W_1 + W_2)(N-1) - 4W_0{}^2)}{2(N+1)W_0}$$

$$Var(C)_R = \frac{W_1(N-1)(N^2 - 3N + 3 - k(N-1)}{W_0 N(N-2)(N-3)} + \frac{N^2 - 3 - k(N-1)^2}{N(N-2)(N-3)} - \frac{(N-1)W_2\ (N^2 + 3N - 6 - k(N^2 - N(N^2 - N + 2))}{4N(N-2)(N-3)W_0}$$

where, all nations are expressed in the equation [**8**].

The value of Geary's *C* ranges from 0 to +2 mean that (0, +1, +2).

When the value of Geary's *C* is zero, it indicates that strong positive spatial autocorrelation and when the value of Geary's *C* is +1 through +2 shows strong negative spatial autocorrelation (Anselin, 1992). If values of any one location are spatially unrelated to any other location, the expected value of *C* will be 1. Due to the squared term in the numerator in Geary's *C* gives greater weight to extreme values than Moran's *I*. As a consequence, the global Moran's *I* is generally preferred in practice (Cliff and Ord, 1981). The null hypothesis of no spatial autocorrelation will be rejected if the calculated value of $|Z(C)| \geq Z_{\alpha/2}$.

3.5.5. Moran Scatter Plot

The Moran scatter plot is a useful visual tool for exploratory spatial analysis because it enables us to assess how similar an observed value is to its neighboring observations. Its horizontal axis is based on the values of the observations and is also known as the response axis. The vertical Y axis is based on the weighted average of the corresponding observation on the horizontal X axis.

The Moran scatter plot provides a visual representation of spatial association (dependence) in the neighborhood around each observation. Depending on their position in the plot, the Moran scatter plot data points express the level of spatial association of each observation with its neighboring ones.

The four different quadrants of the scatter plot correspond to the four types of local spatial association between a region and its neighbors: the first quadrant, (HH) a region with a high value surrounded by regions with high values (top on the right), the second, (LH) a region a with low value surrounded by regions with high values (top on the left), the third (LL) a region with a low value surrounded by regions with low values (bottom on the left) and the last (HL) a region with a high value surrounded by regions with low values (bottom on the right). The first and the third quadrants refer to positive spatial autocorrelation indicating spatial clustering of similar values whereas the second and the forth quadrants represent negative spatial autocorrelation indicating spatial clustering of dissimilar values. The Moran scatter plot may thus be used to visualize typical localizations, i.e. regions in quadrant two or in the quadrant four (Anselin, 1996).

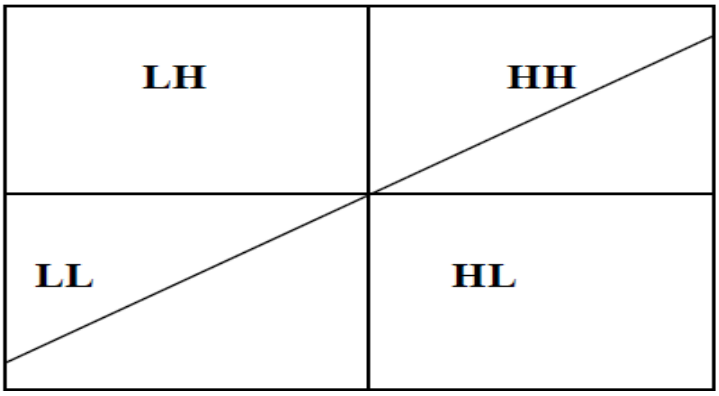


Figure 3.2: Moran's I Scatter Plot

3.5.6. Local Measures of Spatial Autocorrelation

Local spatial autocorrelation is used to identify the regions of significantly high or low value clusters. It includes Local Moran's *I* and Local Ord and Getis' Gi* statistic.

3.5.6.1. Local Moran's *I*

Local Moran's *I* is a local test statistic for spatial autocorrelation, which is used to identify the locations of spatial clusters and spatial outliers.

Local Moran's *I* for each observation measures the extent of significant spatial clustering of similar values around that observation Anselin (1995). The Local Moron's *I* is denoted by I_i and can be given as follows:

$$I_i = \frac{\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(x_i - \bar{x})^2} \qquad [11]$$

The interpretation of values of the local Moran's *I* is analogous to their global counterpart.

For $n \sim \infty$ $Z(I_i)$ follows the standard normal distribution. That means $Z(I_i) \sim N(0,1)$ and the variable x follows normal distribution. Therefore, the Z-statistics of local Moran's *I* is given as follows:

$$Z(I_i) = \frac{I_i - E(I_i)}{sd(I_i)} \qquad [12]$$

where,

$$sd(I_i) = \sqrt{Var(I_i)}$$

$$E(I_i) = \frac{\sum_{i=1}^{N} W_{ij}}{N-1}$$

$$\text{var}(I_i) = W_i \frac{(N-b_2)}{(N-1)} + \frac{2W_i(kh)(2b_2 - N)}{(N-1)(N-2)} - \frac{{W_i}^2}{(N-1)^2}$$

$$b_2 = \frac{m_4}{{m_2}^2},\ m_2 = \sum_{i=1}^{N} \frac{{y_i}^2}{N},\ m_4 = \sum_{i=1}^{N} \frac{{y_i}^4}{N}, \quad W_i = \sum_{i \neq j}^{N} {W_{ij}}^2,\ 2W_{i(kh)} = \sum_{h \neq i}^{N} \sum_{k \neq i}^{N} W_{ik} W_{ih}$$

Here *i*, *k* and *h* represents i^{th}, k^{th} and h^{th} location, respectively.

By the estimating local Moran's *I* scatter plot, we can identify which observations are consistent with the global pattern of positive or negative spatial autocorrelation and which observations run counter to this global pattern (Anselin, 1992).

3.5.6.2. Local Ord and Getis Gi* Statistic

Getis and Ord (1992) introduced a family of statistics, *G*, that can be used as measures of spatial association in a number of circumstances. The local statistics, G_i and $G*i$ enable us to detect pockets of spatial association that may not be evident when using global statistics. Gi and G_i* are used to compare local averages to global averages. The difference between *Gi* and $G*$ are expressed as follows: G_i* statistic includes the value of the point in its calculation, whereas G_i excludes this value and only considers the value of its nearest neighbors (within *d*) against the global average (which also does not include the value at site *i*). G_i* is the more popular of the two statistics because it considers all values within d. In Getis and Ord (1992), the statistic $G_i(d)$ is defined as follows:

$$G_i(d) = \frac{\sum_{j=1}^{N} W_{ij(d)} x_j}{\sum_{j-1}^{N} x_j}, i \neq j \qquad [13]$$

where, $W_{ij}(d)$ is a symmetric one or zero spatial weight matrix with ones for all links defined as being within distance d of a given i; all other links are zero including the link of point i to itself.

The Ord and Getis G_i* test statistic is given as:

$$G^*_i = \frac{\sum_{j=1}^{N} W_{ij} x_j - (\sum W_{ij} + W_{ii})\bar{x}}{S\sqrt{NS^* - \frac{(\sum_{j=1}^{N} W_{ij} + W_{ii})^2}{N-1}}} \qquad [14]$$

where,

W_{ij} is the element of the spatial weights matrix W.

W_{ii} is a weight in the case in which i is in its own neighborhood set.

$\bar{x}$ is the mean of the values on the dependent variable.

$$S^*_i = \sum_{j=1}^{N} W_{ij} + W_{ii}^2$$

S^2 is the sample variance.

The expected values of *Gi** are given by:

$$E(G^*_i) = \frac{\sum_{j=1}^{N} W_{ij}}{N} \qquad [15]$$

where,

N is total number of observations and W_{ij} is expressed in equation [14]

To determine whether or not a particular G_i * score is significant, we will use cluster map and its significance level Z_i.

The variance G_i *is given by:

$$\text{var}(G^*_i) = \frac{W_i^*(N - E(G^*_i)X_{i2}}{N^2(N-1)X_{i1}}$$

where,

$$W^*_i = \sum_{j=1}^{N} W_{ij}$$

$$X_{i1} = \sum_{j=1}^{N} x_j$$

$$X_{i2} = \frac{\sum\sum (X_i X_j)^2}{N} - X_{i1}$$

The Z_i statistics for G^*_i is given by:

$$Z_i = \frac{G^*_i - E(G^*_i)}{\sqrt{\text{var}(G^*_i)}} \qquad [16]$$

The null hypothesis of no spatial clustering will be rejected if the computed value of $|Z_i| \geq Z_{\alpha/2}$. Z score is used extensively in determining confidence thresholds and in assessing statistical significance. Z scores indicate the place of a particular value in a dataset relative to the mean, standardized with respect to the standard deviation. Z = 0 is equivalent to the sample/data mean. Z < 0 is a value less than the mean. Z > 0 is a value greater than the mean.

For statistically significant positive z-scores, the larger the z-score is, the more intense the clustering of high values (hot spot). For statistically significant negative z-scores, the smaller the z-score is, the more intense the clustering of low values (cold spot).For this reason G_i* statistic, unlike Moran's I, cannot distinguish cases of positive spatial autocorrelation from cases of negative spatial autocorrelation (Getis and Ord, 1992).

Local spatial autocorrelation can exist in the absence of global autocorrelation when the clustering at the local level is limited as a proportion of the overall number of observations or when local patterns are off-setting, producing no global pattern as a consequence (Getis and Ord, 1992).

3.6. Diagnostics for Spatial Dependence

As mentioned above the global and local tests of spatial autocorrelation are used to identify the significant spatial dependence. After spatial dependence has been identified the next procedure is to model the spatial autocorrelation through covariates by Moran's *I* and Lagrange Multiplier diagnostics.

Spatial dependence in the error term can be identified by:

- Moran's *I* Diagnostic for Spatial Error Dependence.
- Lagrange Multiplier Diagnostic for Spatial Lag and Spatial Error Dependence.
- Robust Lagrange Multiplier Diagnostics for Spatial Lag and Spatial Error Dependence **.**

3.6.1. The Moran's I Diagnostic for Spatial Error Dependence

The most commonly used specification test for spatial autocorrelation is derived from a statistic developed by Moran (1948) as the two-dimensional analog of a test for univariate time series correlation (Cliff and Ord, 1973). Moran's *I* test has been shown to be locally best invariant (King, 1981) and consistently outperforms other tests in terms of power in simulation experiments (Florax, 1995).

The Moran's *I* Diagnostic test for spatial error autocorrelation is a general test. The Moran's *I* test statistic for spatial error dependence in OLS estimation residuals (Moran, 1950) is given by:

$$I = \frac{N}{S}\frac{e'we}{e'e} \qquad [17]$$

where,

N is the number of observations.

S is the sum of the weights.

e is vector of the residuals from an OLS estimation and W is the spatial weights matrix.

The test statistic *I* (under Ho: λ= 0) is distributed as F-distribution with n and m degree of freedom.

Since,

$$\frac{e'we/S}{e'e/N} = \frac{x^2/n}{x^2/m} \sim F$$

The null hypothesis (Ho) will be rejected if the computed value of $I > F_{\alpha}(n,m)$.

3.6.2. Lagrange Multiplier Diagnostic for Spatial Lag and Spatial Error Dependence

The Lagrange multiplier test for spatial dependence (LM error test) is based on the estimation of the regression model. Anselin and Rey (1991) argue for use of Lagrange Multiplier (LM) diagnostics in OLS specifications. There are two basic LM diagnostics.

The first is a Lagrange Multiplier diagnostic for spatial lag dependence in the presence of covariates in regression model. Spatial dependence in regression models may not only be inflected in the error. Instead it may be accounted by entering a spatial lag WY in the endogenous variable Y. In this case, the regression model reads $Y = X\beta + \varepsilon$ with spatial autocorrelation in the error term as spatially lagged dependent variable $\mu = \rho WY + \varepsilon$.

Under the null hypothesis: $H_0: \rho = 0$, the standard regression model $Y = X\beta + \varepsilon$ holds, while under the alternative hypothesis: $H_1: \rho \neq 0$,

The LM diagnostic for lag dependence test statistic is given by:

$$LM_{lag} = \frac{e'WY \big/ S^2}{N_J} \qquad [18]$$

where,

$$N_J = T(WX\beta)'M(WX\beta) \big/ S^2$$

$$T = tr[(W + W')W]$$

$$S^2 = \frac{e'e}{N}$$

$$M = I - X(X'X)^{-1}X'$$

Here N and e are expressed in the equation [17] and tr is the matrix trace operator, W is the spatial weights matrix for the spatial lagged dependent variable and Y is the value of dependent variable.

The test statistic LM_{lag} under $H_0 : \rho = 0$ is distributed as X^2 (chi-square) with one degree of freedom. The null hypothesis (Ho) will be rejected if the computed value of $LM_{lag} \geq X^2(1, 1\text{-}\alpha)$.

The Second is Lagrange Multiplier diagnostic for spatial error dependence in the presence of covariates in OLS method which is based on the estimation of the regression model $Y = X\beta + \lambda W_V + \varepsilon$ with spatially dependent errors term (v).

The null hypothesis that there is no spatial error dependence:

$Ho: \lambda=0$. This means that OLS estimation of the model $Y = X\beta + \varepsilon$ suffices for conducting the LMerror test. The alternative hypothesis claims a spatial autoregressive coefficient $H_1: \lambda \neq 0$

The test statistic is given as:

$$LM_{error} = \frac{(e'We / S^2)^2}{T} \qquad [19]$$

The test statistic LM error (Under the null Ho: λ=0) is distributed as X^2**(**chi-square**)** with one degree of freedom. The null hypothesis (Ho) will be rejected if the computed value of $LM_{error} > X^2(1, 1-\alpha)$.

For each diagnostic, the null hypothesis is the absence of the particular form of spatial dependence. If the null hypothesis cannot be rejected on either diagnostic, the OLS specification is sufficient to model the spatial dependence estimated via the global and local measures of spatial autocorrelation. If the null hypothesis will be rejected on the Lagrange Multiplier diagnostic for spatial lag dependence, we should proceed by estimating a mixed regressive, spatial autoregressive (spatial lag) model for the spatially lagged dependent variable. Alternatively, if the null hypothesis will be rejected on the Lagrange Multiplier diagnostic for spatial error dependence, we can either proceed with a more fully specified OLS method or a maximum likelihood spatial error specification (Gimpel and Cho, 2004).

3.6.3. Robust Lagrange Multiplier Diagnostics for Spatial Lag and Spatial Error Dependence

The robust Lagrange Multiplier diagnostics for OLS models apply Bera and Yoon's (1996) modified Lagrange Multiplier tests to the diagnosis of spatial lag and spatial error dependence in OLS specifications. The robust Lagrange Multiplier diagnostic for spatial lag dependence in regression model, the Null hypothesis is:$Ho: \rho = 0$

The test statistic for this test problem is:

$$RLM_{lag} = \frac{(e'WY - e'We)^2 \Big/ S^2}{(N_J \rho\beta)^{-1} - T} \qquad [20]$$

where,

$$(N_J \rho\beta)^{-1} = [T + (T + (WX\beta)'M(WX\beta) \Big/ S^2]^{-1}$$

For robust Lagrange Multiplier diagnostic for spatial error dependence in an OLS method, the null hypothesis is: $H_0 : \lambda = 0$ versus the alternative hypothesis is: $H_1 : \lambda \neq 0$.

The test statistics is given by:

$$RML_{error} = \frac{\left[\left(\frac{e'We}{T} - T\right) - (N_J \rho\beta)^{-1}(e'WY)/S^2\right]^2}{\left[T - T^2(N_J \rho\beta)^{-1}\right]} \qquad [21]$$

where all are notations are expressed in [18] and [20] above.

The RML_{error} test statistic under null hypothesis is distributed as X^2 (chi-square) with one degree of freedom. The null hypothesis will be rejected if $RML_{error} > X^2(1,1-\alpha)$ and $LM_{lag} > X^2(1,1-\alpha)$.

The robust Lagrange Multiplier diagnostic for spatial lag (error) dependence tends to reduce power against spatial lag (error) dependence than the unidirectional Lagrange Multiplier diagnostic for spatial lag (error) dependence in the absence of spatial error (lag) dependence. As a result, if the null hypothesis is rejected for the robust LM diagnostic for spatial lag (error) dependence due to the presence of spatial error (lag) dependence, the null hypothesis will also likely be rejected for the non-robust LM diagnostic for spatial lag (error) dependence (Anselin and Rey, 1991). In general, a likelihood ratio test will be employed after estimation to choose the proper spatial regressions specification.

3.7. Modeling Spatial Dependence

Modeling is facilitated with spatial autocorrelation specifications. In the case of spatial data, here spatial dependence is detected; it is very unlikely that the standard hypothesis of uncorrelated observation is true. The usual map analysis tools and the scatter plot can provide the first indications that the observed values are more correlated than would be expected under a condition of independence. In this case, global and local spatial autocorrelation tests on the regression residuals warn of the presence of spatial autocorrelation. If spatial autocorrelation exists, we must specify a model that takes into account the effect of it.

Spatial autoregressive models are the error generating process and operate with spatial weight matrices that specify the strength of interaction between neighboring sites (Cressie, 1993). We used a spatial autoregressive model to measure the relationships between malaria incidence rate and meteorological variables obtained at a neighborhood study area.

There are two ways to incorporate spatial autocorrelation in a spatial autoregressive model (we use the notation presented in Anselin, 1988), depending on where the spatial autoregressive process is believed to occur (Haining, 2003). The first is the spatial lag model, the value of a dependent variable Y at a location is modeled as a function of the independent variables X in that location as well as the values of the dependent variable at the neighboring locations, that is, the spatial lag. A spatial lag is basically the weighted average of the dependent variable values at the neighboring locations (Anselin, 1988).

Spatial lag model assumes that the autoregressive process occurs only in the response variable (lagged-response model), and thus includes a term (ρW) for the spatial autocorrelation in the response variable Y, but also the standard term for the predictors and errors $(X\beta + \varepsilon)$ as used in an ordinary least squares (OLS) regression. The spatial lag model (Cliff and Ord, 1973; Ord, 1975; Bivand, 1984; Anselin, 1988; LeSage and Pace, 2009) is the most frequently encountered specification in spatial econometrics.

The spatial lag model is given by:

$$Y = \rho WY + X\beta + \varepsilon \quad [22]$$

where,

Y is an (N x 1) vector of observations on a dependent variable taken at each of N locations.

X is an (N x k) matrix of exogenous variables.

β is an (k x 1) vector of parameters.

ε is an (N x 1) vector of independent and identically distributed disturbances and ρ is a scalar spatial lag parameter.

The spatial lag model in equation [22] is equivalent to $Y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon$

where,

I is the identity matrix.

ρ is the auto regression parameter.

W is the spatial weights matrix.

β a vector representing the slopes associated with the predictors in the original predictor matrix X and ε *is* the vector of errors. ρWY is a spatial lag term. Spatial lag is essentially a weighted average of the neighboring values of the dependent variable. If the spatial autoregressive parameter ρ is significant, the spatial dependency does exist for the dependent variable. In this case, the spatial lag model can yield a more accurate description of relationship between the dependent variable and independent variables (Anselin, 1998).

The spatial error model addresses the spatial autocorrelation existing in the regression residuals of the OLS methods. The value of the dependent variable Y in a location is redefined as a function of the independent variables *X* and the regression residuals of the neighboring location that is the spatial error. A spatial error is fundamentally a weighted average of the individual residuals of the neighboring locations (Anselin, 1992), which is

added into the model as an additional explanatory variable $Y = X\beta + \lambda W v + \varepsilon$, which is spatial error model.

The spatial error model assumes that the autoregressive process occurs only in the error term and neither in response nor in predictor variables. The model is most similar to the conditional autoregressive model (CAR), with no direction in the error. In this case, the usual OLS method $Y = X\beta+\varepsilon$, is complemented by a term (λWv) which represents the spatial structure (λW) in the spatially dependent error term (v). The spatial error model is expressed by (Cliff and Ord, 1973; Ord, 1975; Ripley, 1981; Anselin, 1988; LeSage and Pace, 2009) and can be given by the following formula.

$$Y = X\beta + \lambda W v + \varepsilon \qquad [23]$$

where λ is a scalar spatial error parameter, and v is a spatially auto correlated disturbance vector with variance and covariance terms specified by a fixed spatial weights matrix and a single coefficient λ.

The matrix $(I - \rho W)^{-1}$ being used in the spatial lag models (Equation (22) incorporates the influence of higher order neighbors. Unlike the Spatial weight matrix W which is a sparse matrix with 0s for all the higher order neighbors, this matrix is no longer sparse as a consequence of the inverse operation. Since most of the elements in this matrix have a non-zero value, the influence of higher order neighbors is implicitly considered in the spatial autoregressive models (LeSage and Pace, 2009). As mentioned earlier, such a spatial weights matrix requires that testing spatial autocorrelation be based upon the division of the study area into regions so that the spatial contiguity of the observations will not be disrupted by a commonly used spatially random selection scheme. Consequently, spatial autoregressive models were fitted using the other dependent variable and indicator variables. When sparse matrix method or approximation are used, motivated by the size of *N*, no analytical asymptotic standard errors for the coefficients in spatial lag.

A. Spatial Autoregressive Model Assumption

The spatial auto regression model has the following assumptions:

- The error terms across different spatial units are correlated with spatial error in OLS regression (the assumption of uncorrelated error terms is violated).
- The dependent variable in a specific location is affected by the independent variables with neighboring locations (the assumption of independent observations is violated).
- All diagonal elements of spatial weight matrix W are zero.
- (I-ρW) and ($I - \lambda W$) are $n \times n$ non-singular matrices.

The independent variables will be tested for heteroscedasticity by using Breusch-Pagan test and Likelihood ratio test for spatial dependence in order to fulfill the basic assumptions of spatial of spatial lag model and spatial error model.

B. The Likelihood Ratio (LR) Test

Several methods are commonly used to compare models. Choice of methods depends on whether the models are nested. If two models are nested, then the reduced model (the one with fewer parameters) is a special case of the more complicated model (the one with more parameters). In other words, if you can convert the complicated model into the reduced one by setting one or more parameters equal to a fixed constant (usually zero), then the models are nested. If there is no way to convert the more complicated model to the reduced one by fixing parameters, then the models are not nested.

The likelihood ratio test is used for nested models. This test is based on the concept of model deviance, which is related to amount of unexplained variation.

Deviance (D) is defined as:

$$D = 2[NLL(E_m) - NLL(E_s)]$$

where NLL (Em) is the negative log likelihood of the model of interest and NLL (Es) is the negative log likelihood of the saturated model. That is a model with as many parameters as observations. Deviances are expected to be chi-squared distributed and a

model can be considered to be "sufficient" if its deviance is consistent with a chi-squared distribution with $df = N - p$, where N is the number of observations and p is the number of model parameters.

If the models are not nested, then they may be compared using information criteria. The two most common information criteria are Akaike's information criterion (AIC) and Bayesian information criterion (BIC).

$$AIC = -2\log(L) + 2p$$
$$BIC = -2\log(L) + p * \log(n)$$

where L is the likelihood, p is the number of parameters in the model, and n is the number of observations. The better model has a lower AIC or BIC. The goal of AIC is to identify the model that is most plausibly generated by the data, and the goal of BIC is to find the best model for prediction where "best" means highest posterior probability. Multiple models can be compared using AIC or BIC, whether they are nested or not. The likelihood Ratio test is denoted by LR test. The LR test statistics for spatial model is given by:

$$LR = [AIC_{ols}] - [AIC_{spatial}] - 2([K_{ols}] - [K_{spatial}]) \quad [24]$$

with $[K_{spatial}] - [K_{ols}]$ degree of freedom.

Where, LR is likelihood value and K is the number of parameters.

If the computed value of $LR > X^2(1, \alpha)$, this indicates a significant spatial lagged dependence and spatial error dependence. This is achieved by having addressed the spatial autocorrelation in the residuals, which might have been caused by the spatial distributions of dependent variable (Anselin, 1996).

3.8. Methods of Parameter Estimation

Maximum likelihood (ML) estimation of spatial lag and spatial error regression models was first outlined by Ord (1975). The point of departure is an assumption of normality for the error terms. The joint likelihood then follows from the multivariate normal distribution for y. Unlike what holds for the classic regression model, the joint log

likelihood for a spatial regression does not equal the sum of the log likelihoods associated with the individual observations. This is due to the two-directional nature of the spatial dependence, which results in a Jacobian term that is the determinant of a full *NxN* matrix.

In the standard linear regression model, spatial dependence can be incorporated in spatial lag and spatial error model. Maximum likelihood method is used to estimate the parameters of the spatial lag and spatial error models by maximizing the probability or likelihood of the sample data.

3.8.1. Maximum Likelihood Spatial Lag Estimation

Spatial lag model is appropriate when the focus of interest is the assessment of the existence and strength of spatial interaction. This is interpreted as substantive spatial dependence in the sense of being directly related to a spatial model.

Ord (1975) gives the maximum likelihood methods for estimating the spatial lag and spatial error models. The logarithm of the determinate of the (*NxN*) asymmetric matrix $(I-\rho W)$ or $(I-\lambda W)$ does not tend to zero, it constraints the parameter values to their feasible range between the inverse of the smallest and largest eigenvalues of W, since for positive autocorrelation, as $\rho \to 1, ln|I-\rho W| \to -\infty$ and analogously for λ. The log likelihood functions for spatial lag models:

$$L(\beta,\rho,\sigma^2) = \frac{-N}{2}\ln(2\pi)\frac{-N}{2}\ln(\sigma^2) + \ln|I-\rho W| - \frac{1}{2\sigma}\left[Y'(I-\rho W)'((I-X(X'X)^{-1}X'(I-\rho W)Y\right]$$

The relationship between the log-determinant term and the sum of squares term in the log likelihood function in the spatial error model is analogous to that in the spatial lag model, but the sum of squares term involves more computation in the case of the spatial error model. In all cases, a simple line search may be used to find ρ or λ, and other coefficients may be calculated using an ancillary regression once this has been done. The general model is more demanding and requires that ρ and λ be found by constrained numerical optimization in two dimensions by searching for the maximum on the surface of the log-

likelihood function, which is like that of the spatial error model with additional terms in $I - \rho W$.

In addition to the above Ord (1975) showed how it can be expressed in function of the eigenvalues ω_i of the spatial weights matrix of $|I - \rho W|$ is equals to $\prod_{i=1}^{N}(1-\rho\omega_i)$. Using this simplification, under the normality assumption, the log-likelihood function for the spatial lag model is given as follows:

$$L_{lag} = \sum_{i=1}^{N}\ln(1-\rho\omega_i) - \frac{-N}{2}\ln(2\pi)\frac{-N}{2}\ln(\sigma^2) - \frac{(Y-\rho WY-X\beta)'(Y-\rho WY-X\beta)}{2\sigma^2} \quad [25]$$

where, ω_i are the eigenvalues of the spatial weights matrix W**.**

The first condition for the ML estimator yield nonlinear (in parameters) equations which are solved by numerical method. The ML estimate of ρ is obtained from a numerical optimization of the concentrated log-likelihood function (Anselin and Bera, 1998) is given as follows:

$$L^{c}{}_{lag} = \frac{-N}{2}\frac{\left[(e*-\rho e_1)'(e*-\rho e_1)\right]}{N} + \sum_{i=1}^{N}\ln(1-\rho\omega_i) \quad [26]$$

Where, $e*$ And e_1 are the residuals from OLS regressions of Y on X and from W*Y* on X, respectively**.** And Y and WY are expressed in equation [22]

Given the maximum likelihood estimate of ρ, the parameters, β, and the error variance, σ^2, are then easily computed. Generally, the estimation of parameters shown as follows. The spatial autoregressive models as defined above given by:

$$Y = \rho WY + X\beta + \varepsilon$$

where,

$$\varepsilon = (I - \rho W)(Y - X\beta)$$

$$\varepsilon = AY - X\beta \quad [27]$$

$$A = (I - \rho W)$$

The joint likelihood of the ε_i is given by (Mead, 1967):

$$L(\varepsilon) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} \exp\left(\frac{-\varepsilon'\varepsilon}{2\sigma^2}\right) \quad [28]$$

However, it is the Y_i that are observed and not the ε_i. Thus it is the joint likelihood of the Y_i that needs to be maximized and not the function given in Equation [28]. From [27] and [28] we have as the joint likelihood function for Y=y is given by:

$$l(y) = |A|\left(\frac{1}{2\sigma^2 2\pi}\right)^{N/2} \exp\left\{-\left(\frac{1}{2\sigma^2}\right)[AY - X\beta]'[AY - X\beta]\right\} \quad [29]$$

Where, |A| is the Jacobian of the transformation from to y**.** The eigenvalues facilitate computation of the Jacobian transformation from an auto correlated to unautocorrelated mathematical space estimation. Let $\omega = \sigma^2$ then, the log-likelihood function is given by:

$$l(y) = \log l(y) = constant - \left(\frac{N}{2}\right)\ln w - \left(\frac{1}{2\omega}\right)(y'A'Ay - 2\beta'X'Ay + \beta'X\beta) + \ln|A| \quad [30]$$

Minimizing l(y) gives the following solutions:

$$\hat{\beta} = (X'X)^{-1}X'Z \quad [31]$$

$$\hat{\omega} = \left(\frac{1}{n}\right)(Z'Z - ZX(X'X)^{-1}X'Z) \quad [32]$$

where,

$$Z = (I - \rho W)y = Ay$$

Let $M = I - X(X'X)^{-1}X'$ be the symmetric and idempotent matrix. Now $\hat{\rho}$ maximizes

$$l(y) = l(y : \hat{\rho}, \hat{\omega}, \hat{\beta}) = constant - \left(\frac{N}{2}\right)\ln\hat{\omega} - \ln|A| \quad [33]$$

Using the simplified expression for $\ln|A|, \hat{\rho}$ minimizes

$$\left(\frac{-2}{n}\right)\sum_{i=}^{N}\ln(1-\rho\omega_i)+\ln\hat{\omega}$$

But

$$\hat{\omega}=\frac{1}{n}Z'MZ$$

$$=\left(\frac{1}{n}\right)y'A'MAY$$

$$=\left(\frac{1}{n}\right)y'(I-\rho W)'M(I-\rho W)y$$

$$=\left(\frac{1}{n}\right)\left[y'My-2\rho y'WMY+\rho^2(Wy)'MWy\right]$$

Thus $\hat{\rho}$ minimizes

$$-\left(\frac{2}{n}\right)\sum_{i=1}^{N}\ln(1-\rho\omega_i)+\ln\left(\left(\frac{1}{n}\right)\left[y'My-2\rho y'WMY+\rho^2(Wy)'MWy\right]\right) \quad [34]$$

The value of ρ that minimizes the function can be obtained by a direct search procedure (Keith, 2010). Finally, the asymptotic variance-covariance matrix for the estimators of the parameters of the mixed endogenous-exogenous procedures is given by:

$$V(\hat{\omega},\hat{\rho},\hat{\beta})=\hat{\omega}\begin{bmatrix} n/2 & \omega\text{tr}(B) & 0' \\ \omega tr(B) & \omega^2 tr(B'B)+\omega\beta'X'B'BX\beta-\alpha\omega^2 & \omega\text{X}'B'\text{X}\beta \\ 0 & \omega\text{X}'B'\text{X}\beta & \omega\text{X}'\text{X} \end{bmatrix}^{-1}$$

where $B=A^{-1}W$ And the rest are expressed in the equation [**18**] and [**25**].

3.8.2. Maximum Likelihood Spatial Error Estimation

Spatial dependence in the regression disturbance term, or a spatial error model is referred to as nuisance dependence. This is appropriate when the concern is with correcting for the potentially biasing influence of the spatial autocorrelation, due to the use of spatial data (irrespective of whether the model of interest is spatial or not).

A spatial error model is a special case of a regression with a non-spherical error term, in which the off-diagonal elements of the covariance matrix express the structure of spatial dependence. Consequently, OLS remains unbiased, but it is no longer efficient and the classical estimators for standard errors will be biased.

The maximum likelihood estimation for the spatial error model employs the error term into log-likelihood function as follows:

$$L_{Error} = \sum_{i=1}^{N} \ln(1-\lambda\omega_i) - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\left(\sigma^2 - \frac{(y-X\beta)^2(I-\lambda W)'(y-X\beta)}{2\sigma^2}\right) \qquad [35]$$

As in the spatial lag model, the ML estimate can also be solved numerically and the estimates are obtained from the optimization of a concentrated log-likelihood function. The concentrated log-likelihood in the parameter, λ, is given by:

$$L^{c}{}_{Error} = \frac{-N}{2}\ln\left[\frac{e'e}{N}\right] + \sum_{i=1}^{N}\ln(1-\lambda\omega_i) \qquad [36]$$

In this formula, e'e is the residual sum of squares from the regression of the spatially filtered variables $Y - \lambda WY$ and $X - \lambda WX$ (Anselin, 1992). The parameters β and error variance σ^2 are then computed, given the maximum likelihood estimate of λ. Both spatial lag and spatial error models are special cases of a more general specification that may include forms of heteroscedasticity as well. This also provides the basis for ML estimation of spatial Seemingly Unrelated Regressions (SUR) models with spatial lag or spatial error terms (Anselin, 1980). Similarly, ML estimation of error components models with spatial lag or spatial error terms can be implemented as well. Spatial models with discrete dependent variables are typically not estimated by means of ML, given the prohibitive nature of evaluating multiple integrals to determine the relevant marginal distributions.

Finally, it is important to note that models with spatial dependence do not fit the classical framework (Rao, 1973) under which the optimal properties (consistency, asymptotic efficiency, and asymptotic normality) of ML estimators are established. This implies that these properties do not necessarily hold and that careful consideration must be given to

the explicit formulation of regularity conditions. In general term aside from the usual restrictions on the variance and higher moments of the model variables, these conditions boil down to constraints on the range of dependence embodied in the spatial weights matrix.

CHAPTER FOUR

RESULTS AND DISCUSSIONS

In this chapter results and discussions for tests of spatial autocorrelation in the malaria incidence rate in order to determine the distribution pattern of malaria, modeling spatial autoregressive model and see the significance and type of relationship that exists between the dependent and independent variables by using Global Moran's , Geary *C* and Local Moran's statistics will be presented. In addition to this, Moran scatter plots are also used to identify local spatial clustering mainly to identify clustering of high values and low values are also presented. The explanatory variables included in this study are expected to have significant effect on the dependent variable malaria case.

Most of the statistical analyses presented in this study were done using SPSS, GeoDa version 1.6 and ArcGIS software. The total numbers of malaria cases in this study are 66,633 with overall malaria incidence rate of 40.69 per 1000 in the year of 2014.

4.1. Spatial Distribution of Malaria Incidence by Woreda

Figure 4.1 shows the spatial distribution of the proportion of malaria incidence in the study area. The highest incidence rate was observed in Begi (86.3) while the lowest was in Mana Sibu (14.49). In general, a higher proportion of malaria incidence was observed in the Western part of the study area while the northern part has low incidence.

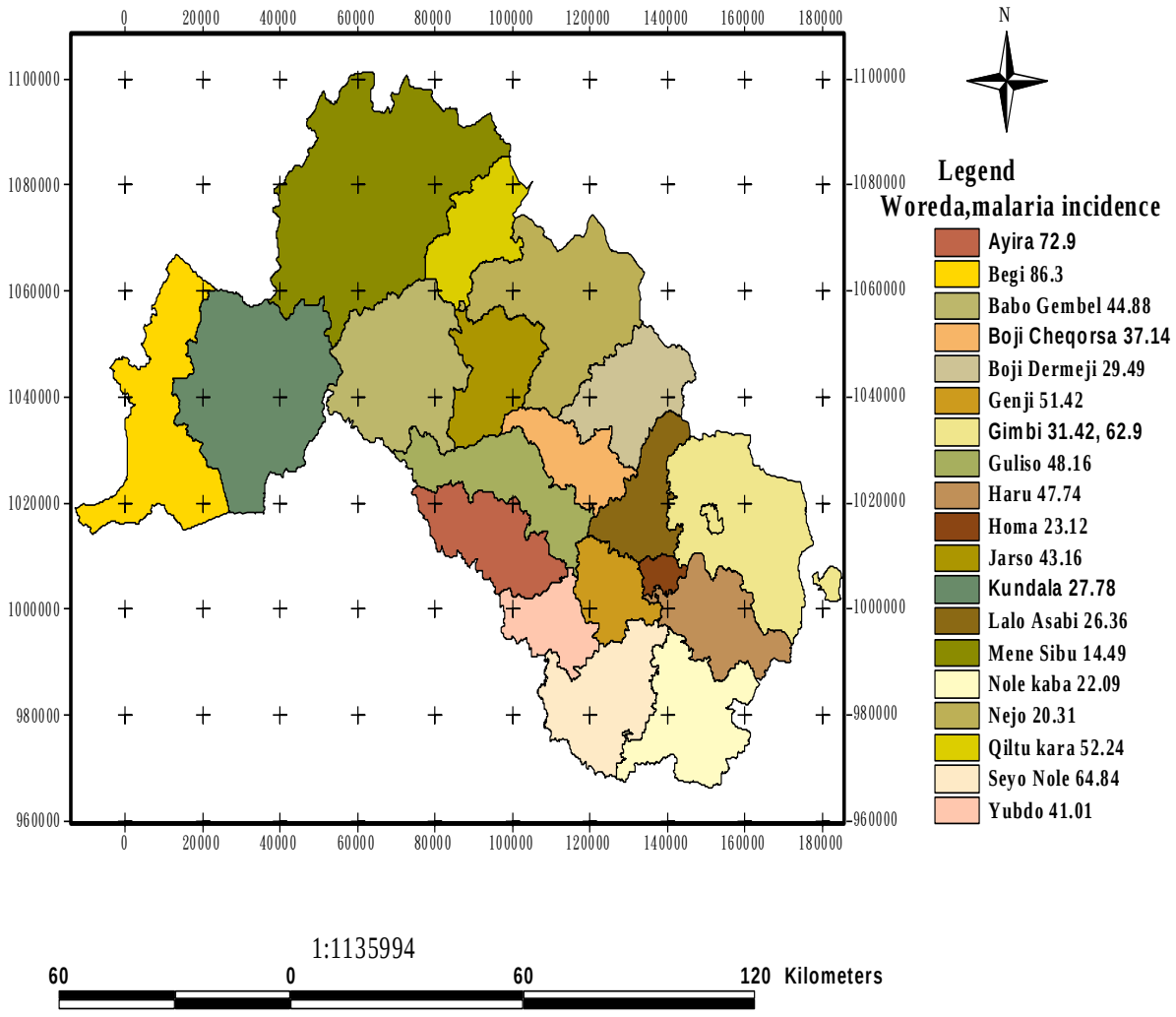


Figure 4.1: Spatial Distribution of the Malaria Incidence rate in Western Wollega Zone, Oromia, Ethiopia

4.2. Tests of Spatial Autocorrelation

The Moran's *I* and Geary's *C* statistic are the most commonly used measure for spatial autocorrelation that shows the degree of spatial association in the datasets. Here the aim was on their application to data analysis, the essential task being to seek for spatial pattern. Spatial autocorrelation analysis includes tests and visualization of both global test

(Moran's *I* and Geary's *C*) and local test for clustering (local Moran's *I* and local Getis and Ord *Gi**) statistics.

The Moran scatter plot is used to visualize Global test in which the slope of the regression line is corresponding to Moran's *I* while Local analysis is based on the local Moran's *I* and local Getis and Ord *Gi** statistic (Anselin, 1995) as discussed in methodological part. Moran's *I* spatial autocorrelation statistic is visualized as the slope in the scatter plot with the spatially lagged variable on the vertical axis and the original variable on the horizontal axis. The variables are standardized to facilitate interpretation and categorization of the type of spatial autocorrelation (cluster or outlier).

4.2.1 Tests for Global Spatial Autocorrelation by using Moran's *I* and Geary's *C* Statistics

Moran's *I* and Geary's *C* Statistics are the global autocorrelation that used for checking whether there is spatial autocorrelation exists over the whole regions. In this section our objective is to identify whether there is positive or negative spatial autocorrelation using Moran's *I* and Geary's *C*.

Positive spatial autocorrelation indicates that woredas are located near to other woredas with similar values, either woredas with high values on the variable being located near to woredas also with high values or the opposite condition (low values nearby other low values). When both Moran *I* and Geary *C* are significant, there is strong evidence that there is significant spatial autocorrelation in the data. If there is negative spatial autocorrelation, which indicates woredas with high values are located near to woredas with low values, or the opposite. The estimating of spatial autocorrelation coefficient is to measure the strength of spatial autocorrelation between neighboring Woreda of malaria incidence as well as to find for spatial pattern or to diagnosis for spatial dependence in regression model and tests are done under the assumption of normality. The results of Moran *I* and Geary's *C* are used for model description. The summary table for the test of Moran *I* and Geary's *C* statistics under normality and randomization are given below respectively.

Table 4.1: summary Output: Results of Global Moran's I and Geary's C Statistics under Normality and Randomization.

Assumption	coefficient	Observed	Expected	Dev. std	Z	Pr>|Z|
Normality	Moran *I*	0.767547	-0.0121	0.1013	7.696	< 0.001
Normality	Geary's *C*	0.374223	0.4832	0.0171	-6.373	< 0.001
Randomization	Moran *I*	0.8853	-0.0455	0.1190	7.82	0.001
Randomization	Geary's *C*	0.0071	0.0617	0.0075	-7.3	0.003

*****significant at 0.05 level

As we can see from Table 4.2, the P-values of the Moran's *I* and Geary's *C* coefficients are less than 0.05 level of significance. Therefore, we can reject the null hypothesis that there is no spatial autocorrelation under normality and independent assumption. In other words, we can accept the alternative hypothesis which shows there is spatial autocorrelation or spatial dependence. In addition to this when the value of Moran's *I* is positive (0.767547) and the value of Geary's *C* (0.374223) statistics is less than one there is the positive spatial autocorrelation or clustering. Since there is significant level of positive global spatial autocorrelation we can statistically identify that there is the presence of malaria incidence in all Woredas given in appendix (Table 1).

Moran's *I* spatial autocorrelation statistic is visualized as the slope in the scatter plot with the spatially lagged variable on the vertical axis and the original variable on the horizontal axis under the assumption of normality. In order to show the distribution of malaria incidence in all woredas we use Moran scatter plot which is given below.

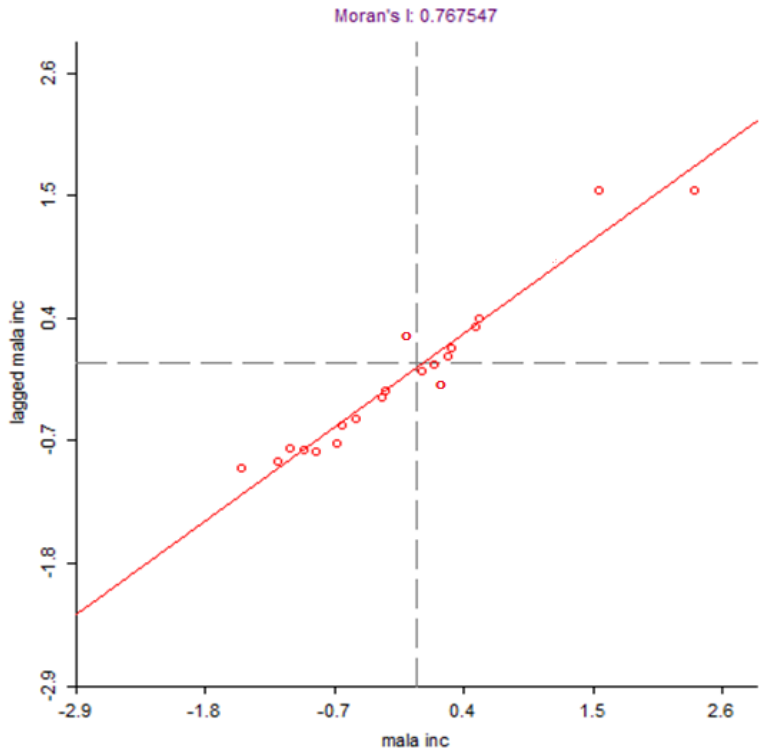


Figure 4.2: Global Moran's I Scatter Plot for Malaria Incidence

As we can see from the Figure 4.2, the malaria incidence cannot be distributed equally in all locations which indicate non-randomization assumptions and also showed that malaria incidence is spatially correlated with neighboring values. Results of Moran's index in the Figure 4.2 showed that there is a significant positive spatial autocorrelation (I = 0.767547) between neighboring Woredas of malaria incidence cases in Western Wollega zone. This means that, the spatial distribution of malaria incidence cases are clustered (globally) in the study area. In this figure the notations **mala inc** is represent malaria incidence while lagged **mala inc** represent lagged malaria incidence, respectively. The Table 4.2 and Figure 4.2 displayed that the presence of malaria incidence in the study area. In order to identify the regions that high or low value clusters local spatial autocorrelations are important.

4.2.2 Tests for local spatial autocorrelation by using Local Moran's *I*

Local spatial Autocorrelation is used to identify the regions of significant high or low value clusters and it includes Local Moran's *I* and Local Ord and Getis' Gi* statistic as

discussed in methodological part. Local spatial autocorrelation can be used as the bases for test on the null hypothesis of no local spatial association. Besides this the Moran scatter plot is one way of identifying local spatial clustering of malaria incidence in Woredas which could be identified as hot spots and cold spots of malaria incidence in study area. To find spatial outliers, we used Moran's Scatter plot and it shows us where values cluster spatially and where values are very different from neighbors (outliers). The results of local Moran *I* as a function of neighboring values at $\alpha = 0.05$ level of significance was given in Table 4.3.

Table 4.2: Summary Output: Results of Local Moran's *I* Test

ID	Woreda	Observed	Expected	Std dev.	Z	P-value
1	Mana Sibu	0.1136	-0.2018	0.1435	2.198	0.0051
2	Lalo Asabi	-0.0133	-0.0019	0.1421	-0.080	0.4230
3	Najo	0.3953	-0.0786	0.1316	3.601	0.0003
4	Ana Gimbi	0.2919	-0.3224	0.2134	2.878	0.0028
5	Kiltukara	0.6431	-0.2340	0.1243	7.056	0.0001
6	Jarso	0.1712	-0.0012	0.0231	7.463	0.0001
7	Homa	0.3841	-0.3125	0.1230	5.684	0.0001
8	Babo Gambel	0.1062	-0.4501	0.2101	2.648	0.0028
9	Ayira	-0.2462	-0.1012	0.41030	-0.353	0.3280
10	Haru	0.4653	-0.0452	0.1201	4.250	0.0002
11	Nole Kaba	0.00243	-0.2156	0.0723	3.016	0.0026
12	Ganji	0.2014	-0.1675	0.0812	4.543	0.0001
13	Guliso	-0.1132	-0.1042	0.6277	-0.014	0.9580
14	Begi	0.5678	-0.2432	0.1435	5.652	0.0001
15	Boji Choqorsa	-0.3098	-0.27657	0.7198	-0.046	0.1340

16	Boji Dirmaji	-0.9324	-0.2435	0.7864	-0.876	0.2370
17	Gimbi Town	0.5101	-0.3147	0.0892	9.247	0.0001
18	Yubdo	0.3064	-0.1723	0.0596	8.032	0.0001
19	Sayo Nole	0.1853	-0.2483	0.0688	6.302	0.0001
20	Kundala	0.2908	-0.1081	0.0377	10.581	0.0001

Table 4.3 revealed that the test results of local Moran *I* as a function of neighboring values which indicates that there is statistically significant local clustering of malaria incidence at 5% level of significance. Statistically significant local clustering of malaria incidence is detected in all the Woredas except in the five Woredas out of a total of 20 Woredas. Those Woredas are Ayira, Guliso, Boji Choqorsa, Boji Dirmaji and Lalo Asabi in which the observed value is less than the expected value. As the result the spatial correlation is negative in these Woredas: this showed that in these Woredas high value is surrounded by low values or low value is surrounded by high values of neighboring Woredas. But in the rest Woredas the observed values are greater than the expected values which indicate that spatial autocorrelation is positive in these Woredas. We can identify hotspot and cold spot of malaria incidence by using local scatter plot based on the results of Local *Gi** test statistics given in the Table 4.4. The local scatter plot of malaria incidence rate was given in the Figure 4.3 that displays Woredas with clustering of similar values and clustering of dissimilar values.

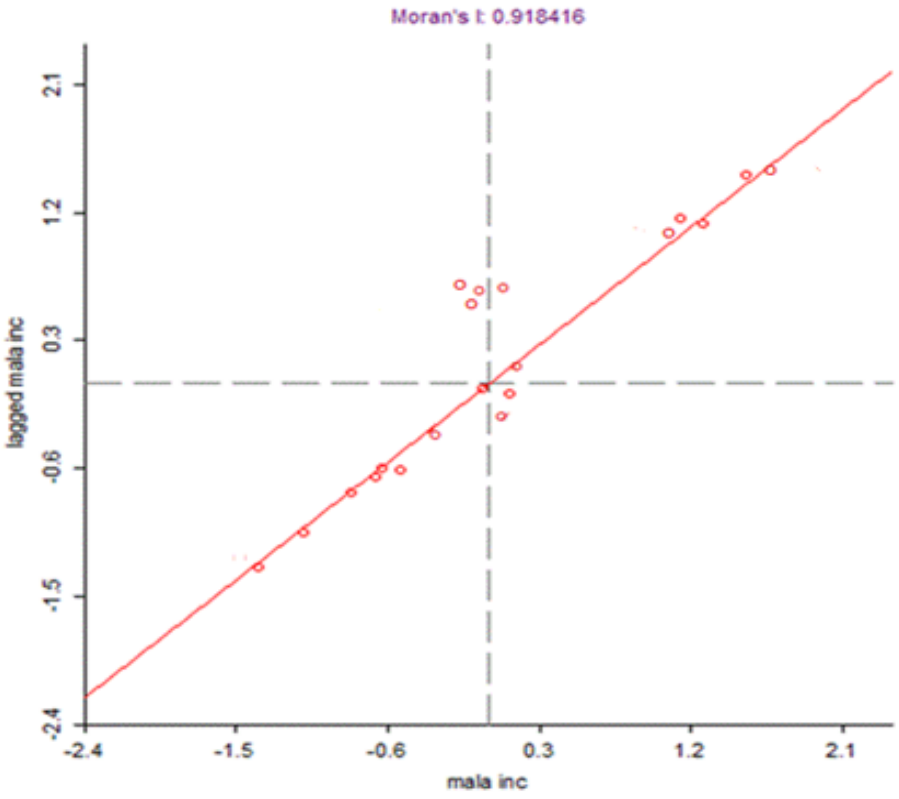


Figure 4.3: Local Moran's *I* Scatter Plot for Malaria Incidence

From the Figure 4.3 Moran scatter plot above, the horizontal axis specifies the observed values of malaria incidence cases while the vertical axis specifies the weighted average of neighboring values of malaria incidence cases. The first and third quadrants indicate the presence of malaria incidence clustering of similar values (positive spatial autocorrelation) while the second and fourth (negative spatial autocorrelation) are used in representing malaria case clustering of dissimilar values**.** Clearly, from the first quadrant (upper right=high-high) of Moran scatter plot we can understand that in seven Woredas, the distribution of malaria incidence are highly clustered. This result indicates that in these seven Woredas, there is high malaria incidence clustering of similar values (hot spots). From the third quadrants (lower left=low-low) we can see that the distribution of malaria incidence in eight Woredas are less clustered. This indicates that in the eight Woredas the distribution of malaria incidence are cold spots. On the other hand, as we can see from the second and fourth quadrant (lower right=high-low and upper left=low-high) of the Moran scatter plot that in five woredas there is malaria incidence clustering

of dissimilar values either high low or low high. Therefore, the first and third quadrants indicated local clustering of malaria incidence, while the second and fourth quadrants did not show local clustering of malaria incidence.

4.2.3. Tests for local spatial autocorrelation by using Local Gi* test statistics

The local Getis and Ord statistic, Gi^*, identified significant local clustering of high (hot spots) or low (cold spots) values of malaria incidence (spatial weighted malaria incidence were standardized by the total malaria incidence at risk in each Woreda) surrounding each Woreda within a nearby location. The spatial weight defined the neighborhood search for each Woreda with nearby locations being expected to have similar values. The observed values were compared with the expected values to indicate if the degree of clustering of malaria cases in the vicinity of a particular Woreda was greater or less than expected by chance.

To correct for multiple comparisons when using G*i, significance levels were adjusted according to Getis and Ord's criteria. The significance of the G*i statistic is assessed by standardized Z value. A positive and significant Z value for the G^*i statistic indicates spatial clustering of high values. A negative and significant Z value for the Gi^* statistic indicates spatial clustering of low values (Getis and Ord's, 1992). This can be identified from the appendix (Figure 1) of Moran's scatter plots of malaria incidence. From these figure malaria incidence is positively spatially correlated with rainfall, minimum temperature, maximum temperature, middle land and low land and negatively spatially correlated with high land respectively. The result of local *Gi** is given in the Table 4.4.

Table 4.3: Summary output: Results of Local Gi* test statistics

ID	Woreda	Observed	Expected	Std dev.	Z	P-value
1	Mana Sibu	0.1765	0.10021	0.03607	2.12	0.0041
2	Lalo Asabi	-0.0240	0.03104	0.110	-5.04	0.2230
3	Najo	0.0428	0.01241	0.01123	2.71	0.0075
4	Ana Gimbi	-0.8460	-0.0215	0.2111	-3.91	0.0004
5	Kiltukara	0.36420	-0.1350	0.1042	4.79	0.0001
6	Jarso	0.10160	-0.0004	0.0401	2.54	0.0031

7	Homa	0.00211	0.41050	0.1150	-3.55	0.0014
8	Babo Gambel	0.0154	0.00115	0.0108	1.319	0.0038
9	Ayira	-0.457	-0.0110	0.3580	-1.244	0.1070
10	Haru	0.02820	0.44010	0.1106	-3.72	0.0012
11	Nole Kaba	0.01301	0.31170	0.1040	-2.87	0.0039
12	Ganji	-0.7326	0.01180	0.1120	-6.91	0.0001
13	Guliso	-0.5612	-0.1136	0.2690	-1.66	0.9580
14	Begi	0.70040	-0.02010	0.1304	5.53	0.0001
15	Boji Choqorsa	0.5450	-0.0110	0.358	1.555	0.060
16	Boji Dirmaji	0.4641	-0.1013	0.2490	2.27	0.2370
17	Gimbi Town	0.01638	0.31470	0.1428	-2.09	0.0060
18	yubdo	-0.8430	0.37675	0.1016	-12.01	0.0001
19	Sayo Nole	-0.01006	0.23018	0.07190	-3.34	0.0012
20	Kundala	0.01108	-0.0804	0.0233	3.93	0.0003

From the Table 4.4 above the standardized Z value of Mana Sibu, Najo, Kiltukara, Jarso, Babo gamble, Kundala and Begi are positive. This indicates that these woredas are spatial clustering of high values and the standardized Z value of Ganji, Yubdo, Ana Gimbi, Gimbi Town, Nole kaba; Sayonole, Haru and Homa are Negative. This showed that these Woredas are spatial clustering of low values. But the rest of Woredas are dissimilar in which high value is surrounded by low value and vice versa. These Woredas are Boj Chokorsa, Boj Dirmaj, Guliso, Ayira and Lalo Asab, respectively. In addition to this Woredas with significant clustering and dissimilar (outliers) can be expressed by using cluster mapping. Cluster mapping helps in classifying issues such as spatial aspects of both internal and external correlations for leading malaria incidence. This is of great aid in assessing spatial risk factors, which in turn facilitates the planning of the most advantageous pattern of malaria distributions and implantations of effective intervention services (Tsai *et al.*, 2009). This study also has practical utility in making the cluster map that can be used to communicate malaria control easily. The cluster map has been used to define a given Woreda within which interventions are scaled and planned according to malaria distribution intensity. This will involve anti-malarial treatment, long-lasting insecticide treated nets (LLINs) and indoor residual spraying (IRS) in the high cluster with neighboring area. The Local Gi* clustering map of malaria incidence was given in the Figure 4.4.

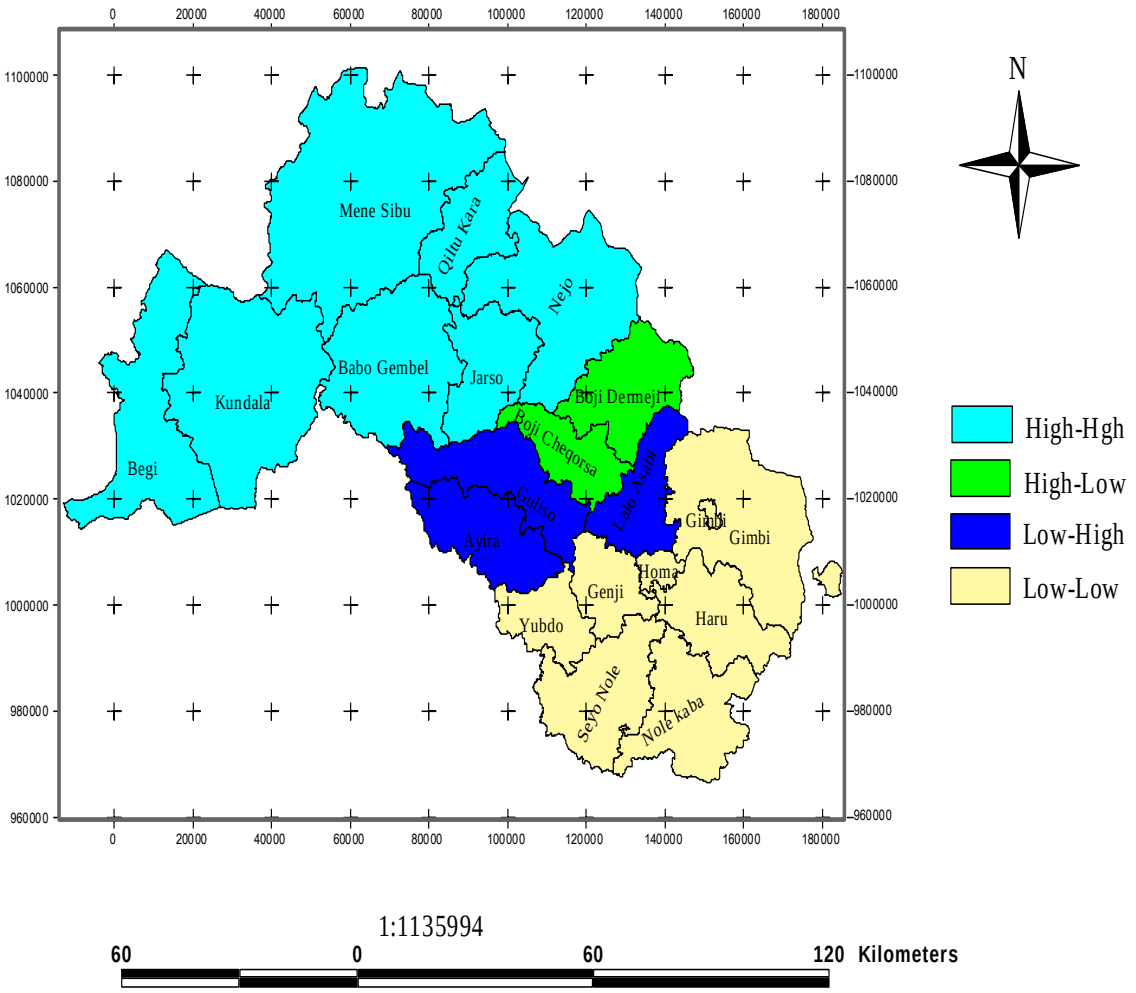


Figure 4.4: Local Gi* Clustering Map of Malaria Incidence in Western Wollega Zone, Oromia, Ethiopia

From the Figure 4.4 it can be observed that Mana Sibu, Kiltukara, Najo, Jarso, Babogambel, Kundala and Begi woredas those have sky blue color are clustered as High-High value and Ana Gimbi, Gimbi Town, Haru, Homa, Nole kaba, Sayo Nole, Ganji and Yubdo woredas those have light yellow are clustered as Low-Low value. But in the case of Boj Dirmaj and Boj Chokorsa those have green color the high value surrounded by

low value that are outliers and Ayira, Guliso and Lalo Asabi woredas those have indigo color the low value surrounded by high value that are also outliers.

4.3. Diagnostic for Spatial Dependence

In this section, the spatial lag model and spatial error model are considered to measure the relationships between malaria incidence and meteorological variables obtained at a neighborhood. The spatial clustering of malaria incidence was indicated by global and local test of spatial autocorrelations. When the data are spatially structured, OLS scores can be biased and their significance be inflated. Therefore, the diagnostic statistic was indicating problem in OLS regression with spatial data. Spatial regression has been used under the assumptions of spatial correlation structures that apply equally across the dataset. The results of diagnostic test that summarizes spatial dependence of the model were given in Table 4.5.

Table 4.4: Summary output: Diagnostic for spatial dependence for weight matrix.

Tests	(Moran's I)/ d.f	Values	prob
Moran's I (error)	0.1424	3.1763	0.00149
Lagrange Multiplier (lag)	1	14.5949	0.00013
Robust LM (lag)	1	18.6159	0.00002
Lagrange (error)	1	1.0380	0.30829
Robust LM (error)	1	2.0591	0.15932

As we can see from Table 4.5, Moran's *I* score of 0.1424 showed that the existence of a strong spatial autocorrelation of the residuals. The Lagrange Multiplier (lag) and Robust LM (lag) both are the simple Lagrange multiplier test for missing spatially lagged dependent variable. The p- values (0.00013 and 0.00002) of Lagrange Multiplier (lag) and Robust LM (lag) were less than 0.05 levels of significance, respectively. Therefore, both Lagrange Multiplier (lag) and Robust LM (lag) are significant that indicate the existence of spatial lag dependence of malaria incidence. However, the Lagrange Multiplier (error) and the Robust Lagrange Multiplier (error) both are the simple Lagrange Multiplier test for spatial error dependence were insignificant because of the p-values are greater than 0.05 level of significance. The importance of robust tests is to

identify the type of spatial dependence. Here, the robust measure for lag is significant and the robust error test is insignificant. These results indicate that less indication for the presence of spatial error model. As the result, the presence of spatial autocorrelation in the data violates the independence assumption of the OLS regression and recommends clear treatment with a spatial autoregressive model.

4.4. Fitting Spatial Autoregressive Models

The spatial regression model integrates spatial autocorrelation effects. The ways of integrating spatial autocorrelation in autoregressive model can be explained as follows: Frist, to model spatial autocorrelation in the error term as a spatially lagged dependent variable as $\mu = \rho WY + \varepsilon$ and the second is λWv which represents the spatial structure (λW) in the spatially dependent error term v, W is the spatial weights matrix characterizing the spatial relationship between every pair of observations, ρ is the spatial autoregressive parameter describing spatial autocorrelation and ε is the independent and normally distributed error term with a constant mean zero and constant variance. As we have seen in diagnostic for spatial dependence Table 4.5 the Lagrange Multiplier (error) and the Robust Lagrange Multiplier (error) for spatial error dependence were insignificant which indicates spatial error model is insignificant.

We can apply spatial lag model $Y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\varepsilon$ in order to fit the model for malaria incidence. The results of spatial Lag Model-Maximum Likelihood Estimation by using GeoDa 1.6 software were given in Table 4.6.

Table 4.5: Results of spatial lag model-maximum likelihood estimation

Variable	Coefficient	Std.Error	Z- value	P-value
W_malaria	0.9652155	0.02379806	40.55858	0.00000
CONSTANT	0.290239	2.107588	1.231399	0.10123
Min Tem	0.527354	0.648287	1.45765	0.01512
Rainfall	0.3879339	0.1247191	3.110462	0.00187

Max Temp	0.783242	0.818433	2.17885	0.02934
Highland	-0.527354	0.648287	-0.75732	0.14512
Middle land	0.1025223	0.04352823	2.355306	0.01851
Low land	0.1499984	0.05148857	2.913236	0.00358

As we can see from Table 4.6 minimum temperature, maximum temperature, rainfall, middle land and lowland are significant since their p- value are less than 0.05 level of significance and they are positively associated with malaria incidence. But highland is insignificant because of the p-value is greater than 0.05 level of significance and it is negatively associated with malaria incidence. This indicated that minimum temperature, maximum temperature, rainfall, middle land and lowland were significant factors which linked to one of the reasons that caused the similarity (clusters) in the malaria distributions and the spatial lag term of malaria incidence represented by W-malaria, seemed as extra indicator. The coefficient parameter ρ of spatial lag reflects the spatially lagged dependence essential in malaria incidence that used to assess the average influence on observations by their neighboring observations and has a positive effect and it is highly significant. Based on this, the equation of fitted spatial lag model is given by:

$Y = 0.290239 + 0.9652155X_1 + 0.27354\ X_2 + 0.3879339X_3 + 0.783242X_4 + 0.1025223X_5 + 0.1499984X_6$

where Y = Malaria incidence, X_1 is the spatially lagged malaria incidence, X_2 is minimum temperature, X_3 is the rainfall, X_4 is maximum temperature, X_5 is middle land, and X_6 is lowland.

From this model, a unit increase in average minimum temperature results in 0.527354 increases in malaria incidence holding the values of other explanatory variables as a constant. A unit increase in rainfall results in 0.3879339 increases in malaria incidence holding the values of the explanatory variables as constant. A unit increase in average maximum temperature results in 0.783242 increases in malaria incidence holding the values of the other explanatory variables as constant. A unit increase of altitude in middle

land results in 0.1025223 increases in malaria incidence by keeping the other explanatory variables as constant and a unit increase of altitude in lowland results in 0.1499984 increase in malaria incidence holding the other explanatory variables as constant. In addition to this, when we compared the values of coefficients of explanatory variables with malaria incidence, the coefficient of maximum temperature (0.783242) was greater than the other explanatory variables. This revealed-that maximum temperature seems to play a more important role in the distribution of the disease than other explanatory variables.

Generally, a rise of minimum temperature in some locations accelerates the distribution dynamics of malaria and maximum temperature would increase the rate of mosquito emergence from breeding places and in the presence of rainfall increased humidity results in longer survival of the vector to transmit the parasite. Therefore, the distribution of malaria incidence within one area is significant associated with variation of minimum temperature, maximum temperature, rainfall, lowland and mid-land zone of neighboring Woredas as determined by standardized spatial weight matrix. This is assumed that malaria incidence in a given Woredas is associated with climatic conditions of the neighboring Woredas. That showed the spatial shift occurs in Tropical regions and temperate regions.

Based on the spatial patterns of residuals analyzed by creating a Moran's *I*, the assumption of spatial autoregressive model can be checked. As mentioned in Table 4.5 the value of the Moran's *I* test statistic for the OLS residuals is 0.1424 and for the lag residuals is 0.965215 in the appendix (Table 3). This indicated that the residuals of the spatial lag models are dependent thereby satisfying the fundamental assumption about the correlation of the error terms. In addition to this about 97.67% in the appendix (Table 3) of malaria incidence are explained by independent variables. The regression diagnostics disclose substantial normality and heteroscedasticity, as well as high spatial autocorrelation. Based on the results of local *Gi** test, we concluded that a spatial lag model is the proper for to indicate relation between malaria incidence and meteorological variables. As discussed in Table 4.5 both LM lag and RLM lag are significant, while both

LM error and RLM error statistic are insignificant which used for the estimation of the spatial lag model discussed in Table 4.6.

The Breusch-Pagan test is used to test the independent variables for heteroscedasticity and Likelihood ratio test for spatial dependence in order to fulfill the basic assumptions of spatial lag model and spatial error model as discussed in methodological part. One of the classical assumptions of the ordinary regression model is that the disturbance variance is constant, or homogeneous, across observations. If this assumption is violated, the errors are said to be heteroscedasticity. The models with spatial dependence do not fit he classical framework under which the optimal properties of ML estimators are established. As we can see from the result of regression diagnostics for spatial lag model given in the Table 4.7, the p-values of Breusch-Pagan test (0.04472) and Likelihood ratio test (0.0000) is less than $\alpha = 0.05$ level of significance which indicates that a significant spatial lagged dependence.

Table 4.6: Summary Output of Regression Diagnostics for Spatial Lag Model

Diagnostics for heteroscedasticity			
Random coefficients			
Test	d.f	Value	p-value
Breusch-Pagan test	6	12.8959	0.04472
Diagnostics for spatial dependence			
Spatial lag dependence for weight matrix			
Test	d.f	Value	p-value
Likelihood ratio test	1	36.1451	0.00000

Significant at 0.05 levels

4.5. Tests for Normality of Residuals

The null hypothesis of the Jarque-Bera test is that the skewness and excess kurtosis are jointly zero. The result of normality test of residuals is given in the Table 4.8.

Table 4.7: Normality Test of Residuals Regression diagnostics

Test on normality of errors			
Test	D.f	Value	Prob
Jarque-Bera	2	0.1803	0.9137

As we can see from Table 4.8 the test of Jarque-Bera statistic is 0.180 with p-value of 0.9137. Since the p-value (0.9137) is greater than $\alpha = 0.05$ level of significance, we accept the null hypothesis that indicates the dataset comes from normal distribution. In addition to this we can understand from the normal plot of malaria incidence which is normal (Figure 2 appendixes).

4.6. Tests for multicollinearity

Multicollinearity is a state of very high intercorrelations among the independent variables and if it present in the data, the statistical inferences made about the data may not be reliable. Multicollinearity caused because of an inaccurate use of dummy variables, the repetition of the same kind of variable, the inclusion of a variable which is computed from other variables in the dataset and when the variables are highly correlated to each other. As the results, the partial regression coefficient due to multicollinearity may not be estimated precisely, the standard errors are likely to be high and a change in the signs as well as in the magnitudes of the partial regression coefficients from one sample to another sample. In addition to this, it is difficult to reject the null hypothesis of any study when multicollinearity is present in the data under study. Multicollinearity can be detected with the help of tolerance and its reciprocal, called variance inflation factor (VIF). If the value of tolerance is less than 0.2 or 0.1 and, simultaneously, the value of VIF 10 and above, then the multicollinearity is problematic. Table 4 in the appendix indicated that there was no series problem with multicollinearity since the tolerance value was greater than 0.2 or 0.1 and variance inflection factor (VIF) was less than 10, respectively.

A bivariate measure of spatial correlation relates the value of a variable at a location to that of a different variable at neighboring locations, as a straightforward generalization of

the concept of spatial autocorrelation. The bivariate Moran Scatter Plot of rainfall, maximum temperature, minimum temperature, highland, middle land and low land with respect to malaria incidence are calculated using neighborhood spatial weight matrix that displayed in the appendix Figure 1. From this Figure the Moran's *I* scatter plots of rainfall (0.105773), maximum temperature (0.50662), minimum temperature (0.523655), highland (-0.325529), middle land (0.121022) and low land (0.64137), respectively. This indicates spatial lagged dependence of malaria incidence is clearly positively spatially correlated rainfall, minimum temperature, maximum temperature low land and middle land. This implies that middle land zone and low land zone areas have higher risk levels of malaria incidence. But Moran's *I* scatter plots of high land was negatively correlated with spatially lagged malaria incidence. Based on these results we suggested that the humidity and higher altitude do have a significant statistical relationship with malaria incidence because of climatic condition varies with elevation. So Woredas with a similar level of humidity and altitude may have conditions that are related to the number of incidence because the rainfall, temperature, low land and middle land zones are significant. Based on this result, the distribution of malaria incidence in woredas with similar humidity and elevation may be similar and the distribution of malaria incidence in woredas with different humidity and elevation may be dissimilar.

4.7. Discussion

Geographical clusters of malaria incidence cases were identified through exploratory spatial data analysis, using Global spatial autocorrelation (Moran's *I* and Geary's *C*) and Local spatial autocorrelation (Local Moran's *I* and Local Ord and Getis' Gi*). This study identified significant spatial clusters of the malaria incidence that was higher or lower in Western Wollega Zone. Having identified spatial clustering in the distribution of malaria cases, the next step was to investigate the underlying individual Woreda and meteorological factors that characterize spatial distribution of malaria incidence. Areas characterized by middle land zone and low land zone were strongly associated with the risk of malaria and risk of spatial clustering and as well as maximum and minimum temperatures were found to be significant that indicating strong relationship between temperature and malaria incidence.

Furthermore, local spatial statistics were used to test the spatial dependency in the patterns of malaria distribution, detect pockets of disease as discussed in the Figure 4.3. Local Moran's *I* Scatter Plot for Malaria Incidence and Figure 4.4: Local Gi* Clustering Map of Malaria Incidence in Western Wollega Zone and identify the relevant spatial scale at which local cluster of malaria occurs. As it is clearly noted from the cluster map figures in seven woredas were hotspot and eight woredas were cold spots and five woredas were dissimilar. To identify woreda with malaria burden and as well as to identify the factors associated with spatial differentials (differences), the malaria incidence distribution was analyzed. The spatial lag model is appropriate when the focus of interest is the assessment of the existence and strength of spatial interaction and it was selected as the appropriate spatial autoregressive model, to account for the spatial autocorrelation (weighted average of the malaria incidence at the neighboring woreda, i.e., spatially lagged malaria incidence). This is interpreted as substantive spatial dependence in the sense of being directly related to a spatial model. From the spatial lag model of maximum likelihood estimation (Table 4.6) we can understood that there were statistically significant between malaria incidences and rainfall, minimum temperature, maximum temperature, middle land and low land areas at $\alpha = 0.05$ level of significance. But there were statistically insignificant between malaria incidence and high land area at $\alpha = 0.05$ level of significance. Asnakew analyzed malaria clustering in East Wollega by using global spatial autocorrelation and local spatial autocorrelation and statistical spatial analysis of malaria incidence by age, temperature and village through time revealed the presence of significant spatio-temporal variations. Yeshiwondimet examined the global and local patterns of malaria distribution in 543 villages in Ethiopia using individual-level morbidity data collected from six laboratories and reported that malaria incidence varies according to gender and age with age less five years and above showing a statistically significant malaria incidence.

Chapter Five

Conclusions and Recommendations

This study was significance to provide vital information which is useful in fighting and reducing malaria transmission through mosquito bites and describes the spatial pattern of malaria distribution in Western Wollega Zone using routinely collected individual patient morbidity from health care facilities and meteorological data. These sections give a synthesis about the output of this study in line with the objectives and with an emphasis on the verification of the hypothesis and finally, recommendations are addressed to researchers and local authorities.

5.1. Conclusions

The results of this study showed that the incidence of malaria in Western Wollega Zone displays a spatial pattern which is dependent on some meteorological variables. From this study, we concluded that the spatial pattern of malaria incidence is clustered rather than randomly distributed. There is clear evidence as the incidence of malaria in the study area is significantly clustered indicating high levels in the northern and western part of the zone and low levels in southern and eastern part of the Zone. In other words, it is cogently dissimilar in the central parts of the study area.

From the result of the study, it is concluded that in global Moran's *I* and Geary's C test statistic indicates significant clustering (spatial autocorrelation) among neighboring Woredas in the study area. A significant local clustering of malaria incidence occurs among Woredas within neighboring woredas. As results local test statistic suggested that there was significant clustering of malaria incidence. The results of the model showed those local risk factors such as maximum temperature, minimum temperature, rainfall, highland, middle land and low land as explained by spatial lag model might all be important in explaining the observed local clustering of malaria incidence as a whole.

5.2. Recommendations

This study has attempted to analyze the spatial distribution of malaria incidence in the Western Wollega Zone. The results of this study revealed that in the study area malaria

incidence case loading pattern varies from woreda to woreda and also clustered. The presence of spatial dependence between woredas was also recognized. Based on the results obtained from the fitted models the following recommendations can be given to researchers and to local authorities (concerned body) as follows:

A. **To researcher's**

- Malaria clusters were identified in this study and only differences among the three main clusters (hot spot, cold spot and dissimilar or not significant) were analyzed. There is a need to do more in depth study analyzing causes underlying each of the detecting hot spots in order to identify specific intervention measures.
- This study trusted on the incidence data assuming the presence of the vector. As incidence data are not only collected in settlement areas, for planning purposes, the interpolation of results will not accurately reach the whole Oromia zone. There is a need to do further research for example using Species Distribution modeling, based on anopheles mosquito presence data, which can be collected for the whole Oromia zone.

B. **To Local authorities (concerned bodies)**

- Based on this results the interventions should be facilitated in highly clustered malaria distribution (hot spot) areas by giving special attention in affecting intervention and health services to the highly risk exposed woredas and neighboring woredas.
- Provide concentrated family advice especially in woredas identified as highly clustered malaria distribution (hot spots).

References

1. Abdullah, S. (2010). Patterns of age-specific mortality in children in endemic areas of sub-Saharan Africa. Am J Trop Med Hyg, **77**(6): 99-105.
2. Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, 20: 8-17.
3. Anselin, L. (1988). Spatial econometrics: Methods and models. New York, New York.
4. Anselin, L. (1992). "SpaceStat Tutorial: A Workbook for Using SpaceStat in the analysis of Spatial Data". Typescript. University of Illinois at Urbana Champaign, 8-67.
5. Anselin, L. (1995). Local indicators of spatial association - LISA. *Geographical.*
6. Anselin, L. (1996). The Moran scatter plot as an exploratory spatial data analysis tool to assess local instability in spatial association, 111-125.
7. Anselin, L. (1998). ''Interactive Techniques and Exploratory Spatial Data Analysis.'' In Geographical Information Systems: Principles, Techniques, Management and Applications, *251–64.*
8. Anselin, L. and Bera, K. (1998). Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics, *Handbook of Applied Economic Statistics* eds. Aman Ullah and David E.A.Giles. New York: Marcel Dekker.
9. Anselin, L. and Rey, S. (1991). "Properties of Tests for Spatial Dependence in Linear Regression Models" *Geographical Analysis,* **23**(2): 11-34.
10. Asnakew, K., Yeshiwondim, M., Sucharita, G., Afework, T. and Dereje, O. (2007). Spatial analysis of malaria incidence at the village level in areas with unstable transmission in Ethiopia. *International Journal of Health Geographics,* **8**:1-11.
11. Auto, A., Jalkanen, S. & Rovainen, A. (2013). Nuclear imaging of inflammantation homing associated molecules as targets *EJMMI Res* **3**: 1.
12. Awash, T., Ingrid, P. and Borrel, N. (2009)**.** A Temporal-Spatial Analysis of Malaria Transmission in Adama (Nazereth), Ethiopia. *American Journal of Tropical Medicine and Hyg,* **81** (6): 944-949.

13. Brooker, S. Clarke, J. K. Njagi et al.(2004), Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of Cameron. Tropical Medicine and International Health, **757–766.**
14. Brown, D. (1995). Spatial statistics and GIS applied to internal migration in Rwanda, Central Africa. In: S.L. Arlinghaus and D.A. Griffith (Eds), Practical handbook of spatial statistics. Boca Raton, Florida: CRC Press.
15. Cliff, A. and J.K. Ord (1972). Testing for spatial autocorrelation among regression residuals, *Geographic Analysis*, **4**:267-284.
16. Cliff, A. and J.K.Ord (1981). Spatial processes models and applications. (London, Pion).
17. Cressie, N. (1993). Statistics for spatial data. Revised Edition. John Wiley & Sons, Inc.New York.
18. Fanello, C., Karema, C., Van Dorem, W., Overmeir, C., Ngamije, D., and Alessandr U. (2007). A randomized trial to assess the safety and efficacy of artemether Immefantine (Cortem) for the treatment of uncomplicated plasmodium falciparum malaria in Rwanda. *Transmission of the royal society of Tropical Medicine and hygiene*, **101**(4): 344-350.
19. Geary, R. (1954). The contiguity ratio and statistical mapping. The Incorporated Statistical, **5**: 115.
20. Getis, A. and Ord, J. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, **24**(3):189-206.
21. Grillet, M., Roberto, B. and Marie, J. (2010): Disentangling the Effect of Local and Global Spatial Variation on a Mosquito-Borne Infection in a Neotropical Environment, *Am. Journal. Trop. Med. Hyg*, **82**(2): 1-8.
22. Hay, S. I., Guerra, C. A., Gething, P. W., Patil, A. P., Tatem, A. J., Noor, A. M., Snow, R. W., 2011. A world malaria map: Plasmodium falciparum endemicity in 2007. PLoS Med, **6**(3).
23. Ingrid, P., Luisa, N., Wafaa, E., and Awash, T. **(**2009**).** A Temporal-Spatial Analysis of Malaria Transmission in Adama, Ethiopia, *the American Journal of Tropical Medicine and Hygiene*, **86**(6): 3-4.

24. Jarque, C. and Bera, K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters,* **6**: 255-259.
25. Jeong, W., and Gluck, M. (2002). Spatial maps with auditor and haptic display. Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan.
26. Kerema,C., Aregawi, A., Kabayizat, A, Mulindahabi, M., Fall, I., Binagwaho, A., (2012). Trends in malaria cases, Hospital admissions and death following scale-up of ant- malarial-interventions, Rwanda, *Malaria journal,* **11**(1): 236.
27. LeSage, J. (1998). "Bayesian Estimation of Spatial Autoregressive Models." *International Regional Science Review,* **20**: 115-126.
28. LeSage, J., and Pace, K. (2009). Introduction to spatial econometrics. Boca Raton, Florida: CRC Press.
29. Lindsay SW, Bodker R, Malima R, Msangeni HA, Kisinza W. 2013. Effect of 2009-2010 El Nino on highland malaria in Tanzania. Lancet. : **355**: 989-90.
30. Machaut, V., Vignolles, C., Borchi, F., Vauntson, F., Briolant,S., Rogier, C. (2011). The use of remotely sensed environmental data in the study of malaria. Geospatial health, **5**(2): 151-168.
31. Matthew, L. (2005). The Utility of Geographical Information Systems (GIS) and Spatial temporal cluster Analysis in Tuberculosis Surveillance in Harris, Texas. *International Journal of Health Geographics,* **24**: 3-25.
32. Mbogo, C. M., Mwangangi, J. M., Nzovu, J., Gu, W., Yan, G., Gunter, J. T., Beier, J. C., 2003. Spatial and temporal heterogeneity of Anopheles mosquitoes and Plasmodium falciparum transmission along the Kenyan coast. The American Journal of Tropical Medicine and Hygiene, **68**(6):734-742.
33. Mead, R., (1967). A Mathematical Model for the Estimation of Interplant Competition, *Biometrics,* **23**: 190-205.
34. Moran, P. (1950). Notes on continuous stochastic phenomena. *Biometrika,* **37**:17-23.
35. Mwangangi, J, Ephantus J. Muturi, Simon, M., Muriu, Joseph, N., Midega, J. and Mbogo, C. (2011). The role of Anopheles arabiensis and coustan in indoor malaria

and outdoor malaria transmission in Taveta Distinct, Kenya. *Parasites & vector*, **6**: 114.

36. NISR. (2012). Rwanda Demographic and Health Survey 2010, Kigali, Rwanda, pp 411.
37. Ord, J. and Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis*, **27** (4): 286-306.
38. Plantinga, A., Helvoigt, T., and Walker, K. (2014). Critical habitat for threatened and endangered Species: How should economic costs be evaluated? *Journals of Environmental Management*, **130**(2), 127-135.
39. President's Malaria Initiative (2010). Malaria Operational Plan, Ethiopia.
40. Sachs, J. and P. Malaney (2011). *The economic and social burden of malaria*, **415**(6872): 680-5.
41. Shillu (2003) Association between land cover and habitat productivity of malaria vectors in Eritrea.
42. Stevens, K., Pfeiffer, D. (2011). Spatial modeling of disease using data-and knowledge-driven approaches. *Spatial and Spatio-temporal Epidemiology*, **2**(3): 125-133.
43. Tobler, W. (1979). Cellular geography. In S. Gale and G. Olsson (Eds.), Philosophy in Geography, Dordrecht, Reidel (pp. 379).
44. Tsai**,** P., Men-Lung, L., Chien-Min, C. and Cheng-Hwang, P. (2009). Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006; *BioMed Central Ltd*.
45. WHO (2012).World Malaria Report Geneva, World health Organization.
46. WHO (2010). World Malaria Report *2010, WHO Global Malaria Programme* Geneva, World Health Organization (pp.137).
47. Yeshiwondem, A., Sucharita, G.**,** Afework, T. and Hrishikesh, P. (2009). Spatial analysis of malaria incidence at the village level in areas with unstable transmission in Ethiopia, *International Journal of Health Geographics*, *vol. 8*, can be found online at: http://www.ij- healthgeographics.com/content/8/1/5.

48. Zayeri, F., Salehi, M. and Pirhosseine, H. (2011). Geographical mapping and Bayesian spatial modeling of malaria incidence in Sistan and Baluchistan province, Iran, Asia pacific journals of Tropical medicine, **4**(12): 958-992.
49. Zhou, G., Munga, S., Minakawa, N., Githeko, A. K., and Yan, G., 2011. Spatial Relationship between Adult Malaria Vector Abundance and Environmental Factors in Uganda Highlands. The American Journal of Tropical Medicine and Hygiene, **77**(1):29-35.

Appendix

Table 1: Malaria Incidence per 1000 people in each Woreda in Western Wollega Zone, Ethiopia

No	Woredas	Projected population size (2014) X	Malaria case (2014) Y	Malaria incidence rate= $\frac{Y}{X} * 1000$
1	ManaSibu	152,958	2216	14.49
2	Najo	160,927	3269	20.31
3	Anagimbi	88,788	2883	32.49
4	Lalo Asabi	92,123	2428	26.36
5	Kiltukara	62,738	3340	52.24
6	Boji Dirmaji	52,484	1548	29.49
7	Guliso	84,773	4000	47.18
8	Jarso	58,546	2520	43.04
9	Kundala	115,172	3493	30.32
10	Boji Choqorsa	58,459	2187	37.14
11	Babo Gambel	72,778	3266	44.88
12	Yubdo	46,725	1716	36.72
13	Ganji	71,852	3698	51.47
14	Homa	81,057	1974	24.35
15	Nole kaba	72,258	1596	22.09
16	Begi	144,111	12437	86.30
17	Sayonole	90,932	5796	63.74
18	Haru	29,742	1420	47.74
19	Ayira	57,843	4116	71.15
20	Gimbi (Town)	43,397	2730	62.9
	Total	1,637,663	66,633	40.69

Construction of weight matrix based on Queen's method

Table 2: Weight Matrix.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
3	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
5	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
6	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
9	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
10	0	0	0	1	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
12	0	0	0	0	0	0	0	1	1	1	0	0	1	0	1	1	0	0	0	1
13	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	1	1	1	1	1
14	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	1	1	1	1
15	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	1
16	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1
17	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1
18	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1
19	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1
20	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Woreda	Mana Sibu	Najo	Kiltu Kara	Jarso	Babo Gambel	Kundala	Begi	Boji Chokorsa	Boji Dirmaji	Guliso	Ayira	Lalo Asabi	Ganji	Yubdo	Ana Gimbi	Gimbi Town	Nole Kaba	Sayo Nole	Haru	Homa

Table3: Results of Maximum Likelihood Estimation (spatial lag model)

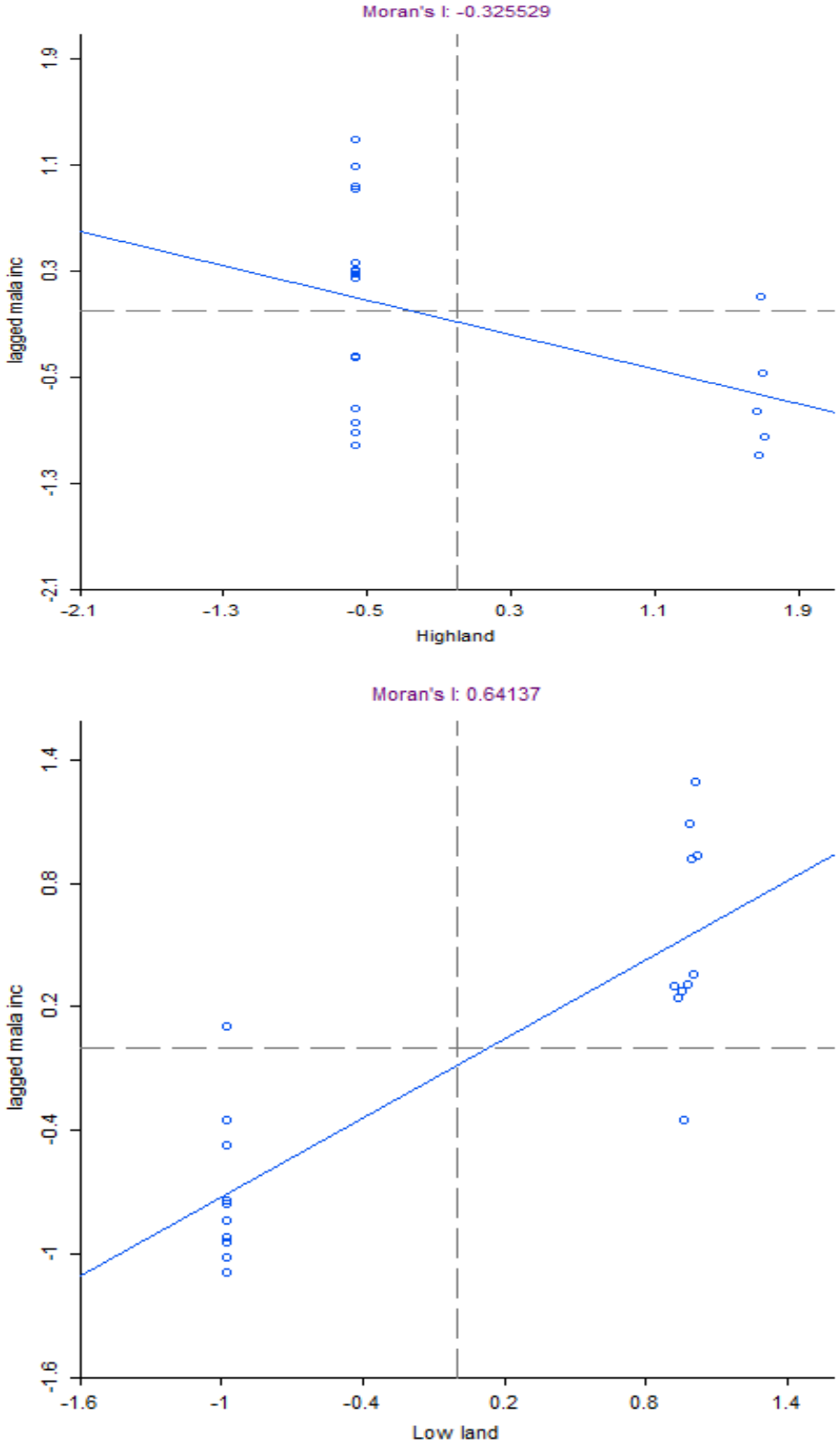
Dependent Variable	: malaria incidence	Number of Observations	: 20
S.D. dependent var	: 18.5293	Degrees of Freedom	: 12
Lag coeff. (Rho)	: 0.965215		
R-squared	: 0.976777	Log likelihood	: -53.4648
Sq. Correlation	: +	Akaike info criterion	: 122.93
Sigma-square	: 7.97347	Schwarz criterion	: 130.895
S.E of regression	: 2.82373		

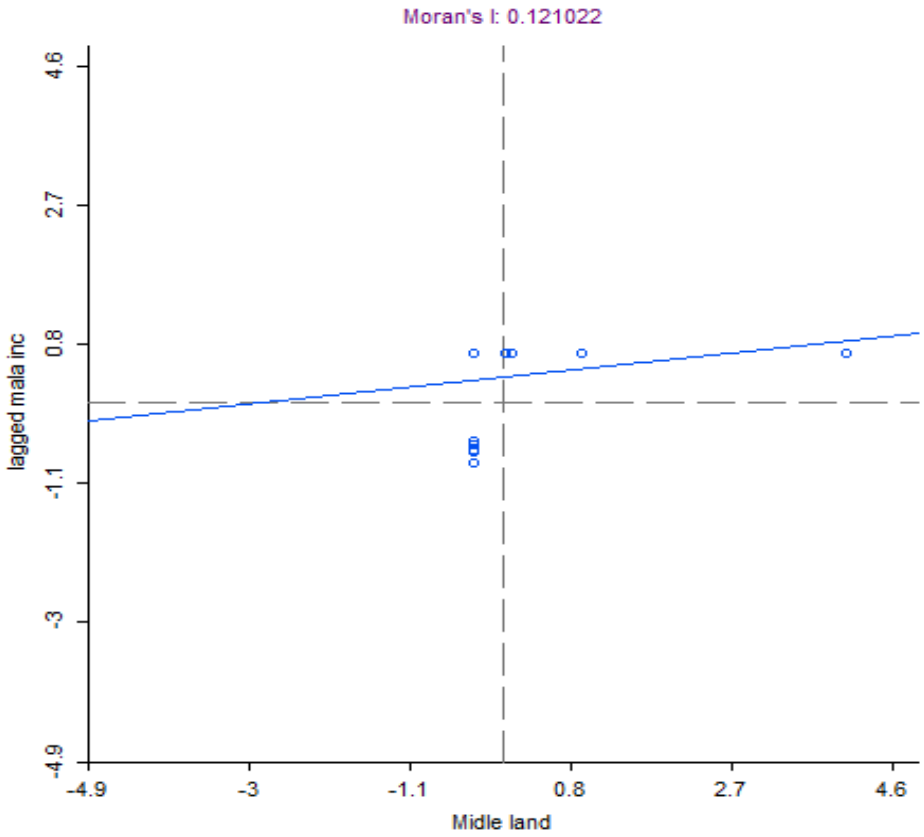
Table 4: Diagnostic for Multicollinearity

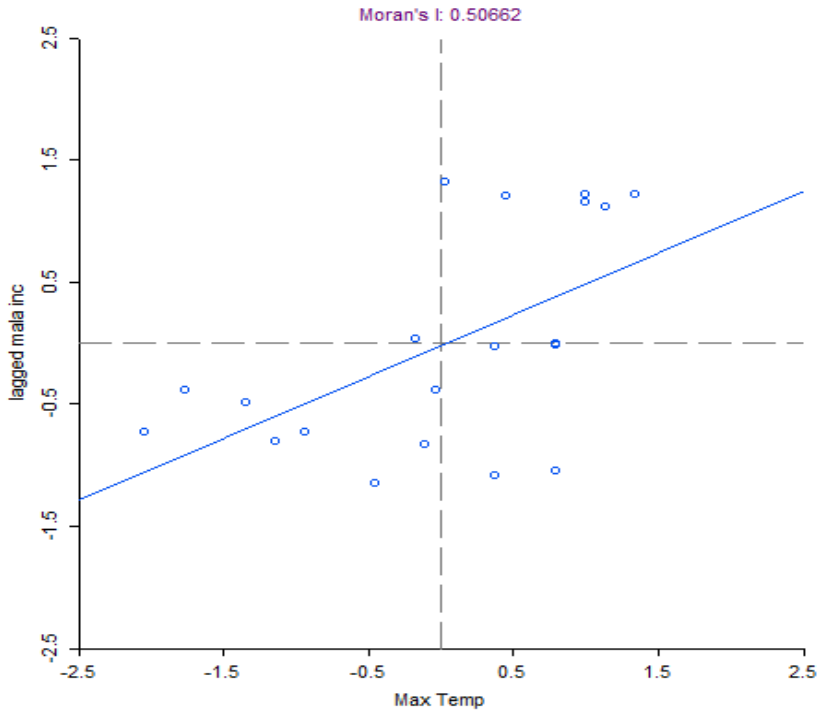
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	133.765	89.277		1.498	.006		
Rainfall	.657	.651	.238	1.010	.003	.933	1.071
Min Tem	.056	1.779	.543	1.156	.001	.236	4.237
Max Temp	.410	1.404	.397	1.004	.001	.334	2.998
Highland	-.005	.006	-.302	-.860	.000	.423	2.363
Low land	.008	.007	.396	1.122	.001	.417	2.399

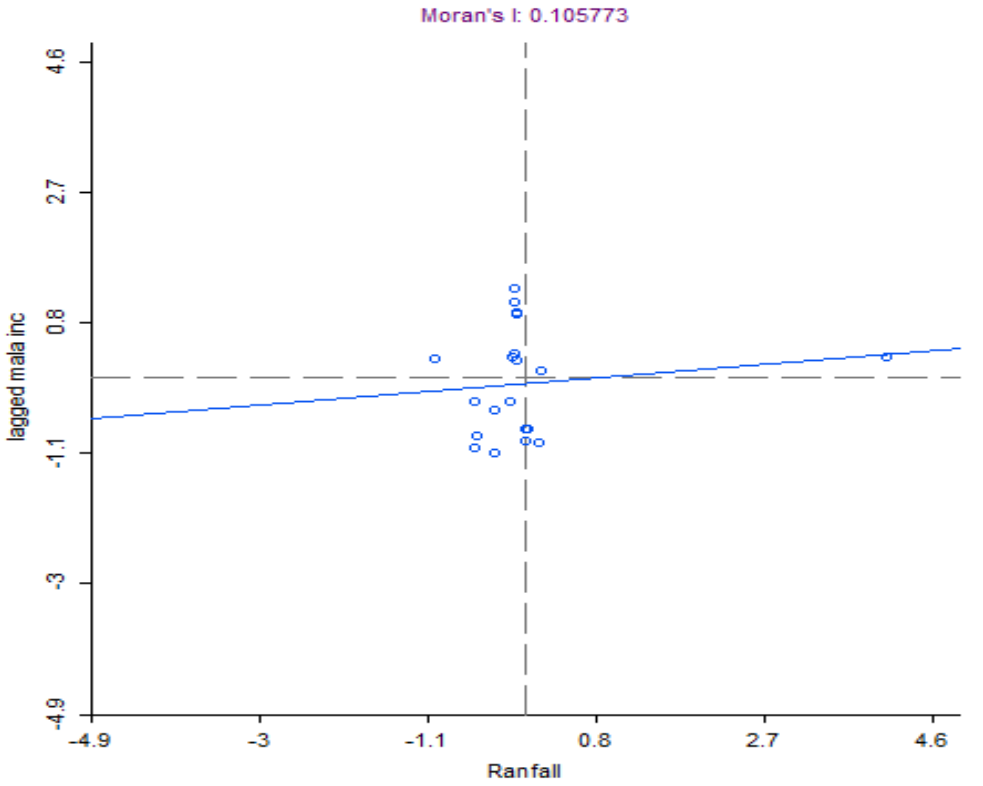
a. Dependent Variable: malaria incidence

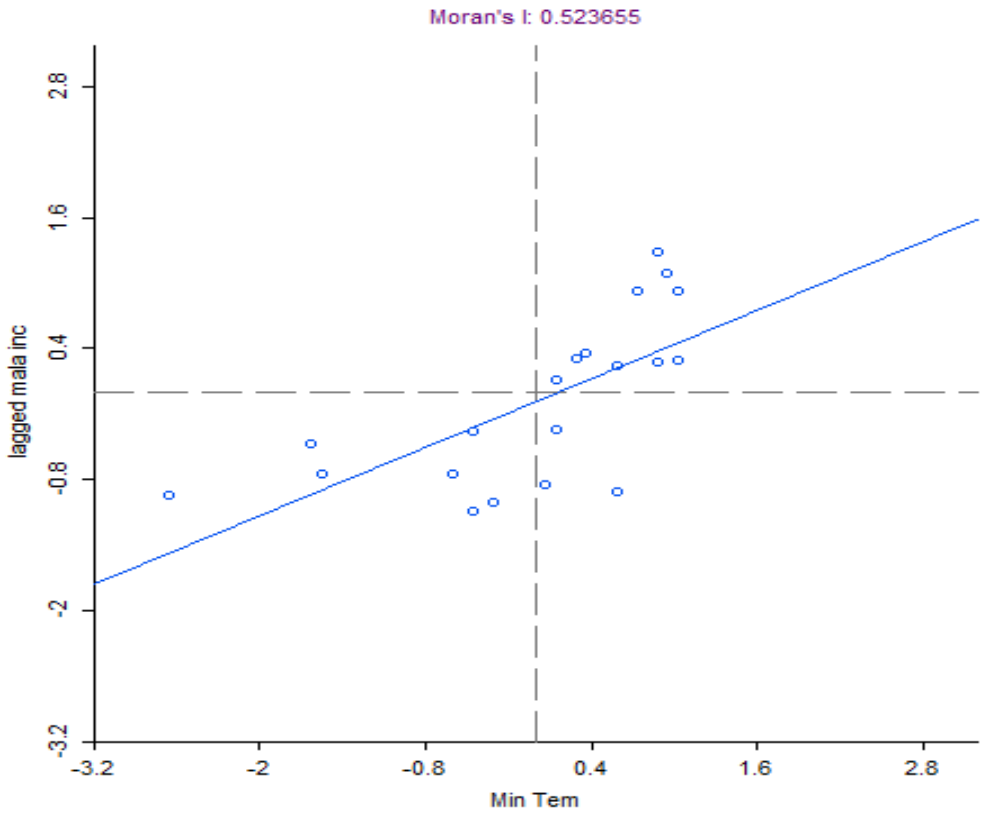
Figure 1: Bivariate Moran's Scatter Plot Based on Neighborhood




Moran's I: 0.121022
lagged mala inc
4.6
2.7
0.8
-1.1
-3
-4.9
-4.9
-3
-1.1
0.8
2.7
4.6
Midle land


Moran's I: 0.50662
lagged mala inc
2.5
1.5
0.5
-0.5
-1.5
-2.5
-2.5
-1.5
-0.5
0.5
1.5
2.5
Max Temp


Moran's I: 0.105773
lagged mala inc
4.6
2.7
0.8
-1.1
-3
-4.9
-4.9
-3
-1.1
0.8
2.7
4.6
Ranfall


Moran's I: 0.523655
lagged mala inc
2.8
1.6
0.4
-0.8
-2
-3.2
-3.2
-2
-0.8
0.4
1.6
2.8
Min Tem

Note: Lagged mala Inc = Lagged malaria incidence

Max Temp = Maximum Temperature

Min Tem = Minimum Temperature

Figure 2: Normal Plot